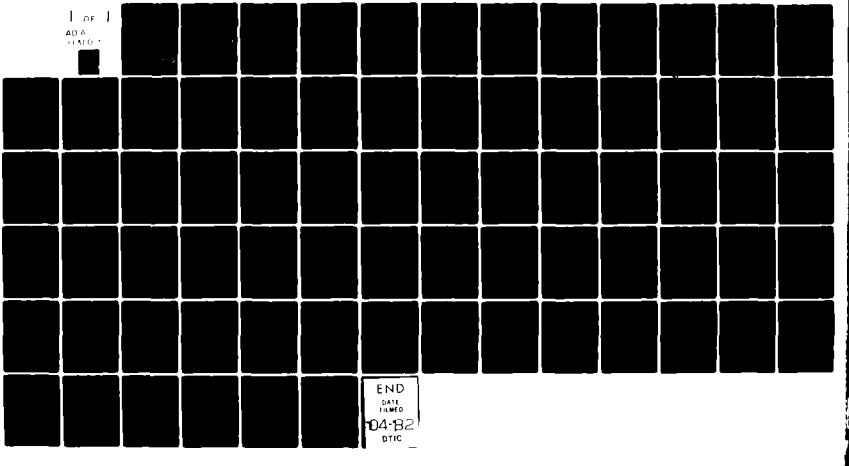
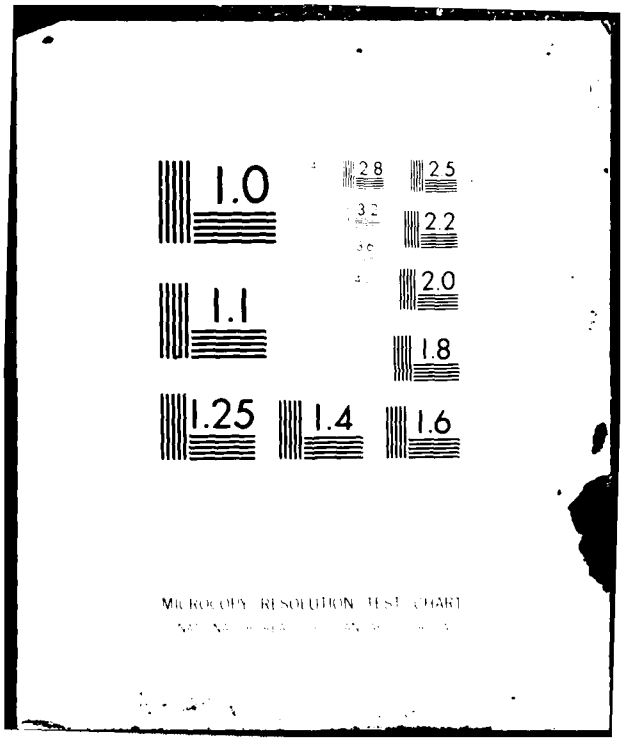
AD-A113 107 SYSTEMS AND APPLIED SCIENCES CORP RIVERDALE MD F/G 4/2
ESTIMATING THE TOPS, BASES, AND AMOUNT OF CLOUDINESS FROM IN SI--ETC(U)
NOV 81 C N TOUART, R SHAPIRO, P J MANSFIELD F19628-81-C-0017
UNCLASSIFIED AFGL-TR-81-0351 NL




1.0
2.8
2.5
3.2
2.2
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART

(12)

AFGL-TR-81-0351

ESTIMATING THE TOPS, BASES, AND AMOUNT OF CLOUDINESS FROM IN SITU SAMPLING

C.N. Touart
Ralph Shapiro
Peter J. Mansfield
Randy Schechter

Systems and Applied Sciences Corporation
6811 Kenilworth Avenue
P. O. Box 308
Riverdale, Maryland 20737

November 17, 1981

Final Report for Period November 1, 1980 - November 17, 1981

Approved for public release; distribution unlimited

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANSCOM AFB, MASSACHUSETTS 01731

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFGL-TR-81-0351	2. GOVT ACCESSION NO. AD A113 127	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ESTIMATING THE TOPS, BASES, AND AMOUNT OF CLOUDINESS FROM IN SITU SAMPLING		5. TYPE OF REPORT & PERIOD COVERED Final Report 1 Nov 1980 - 17 Nov 1981
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) C. N. Touart Randy Schechter Ralph Shapiro Peter J. Mansfield		8. CONTRACT OR GRANT NUMBER(s) F19628-81-C-0017
9. PERFORMING ORGANIZATION NAME AND ADDRESS Systems and Applied Sciences Corporation 6811 Kenilworth Avenue, P.O. Box 308 Riverdale, MD 20737		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 63707F 268801AE
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Geophysics Laboratory Hanscom AFB, Massachusetts 01731 Contract Manager/ F. J. Brousaides/LYS		12. REPORT DATE November 17, 1981
		13. NUMBER OF PAGES 73
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

CLOUDINESS	CLOUD TOP
MEASUREMENT	CLOUD AMOUNT
CLOUD BASE	AUTOCORRELATION

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

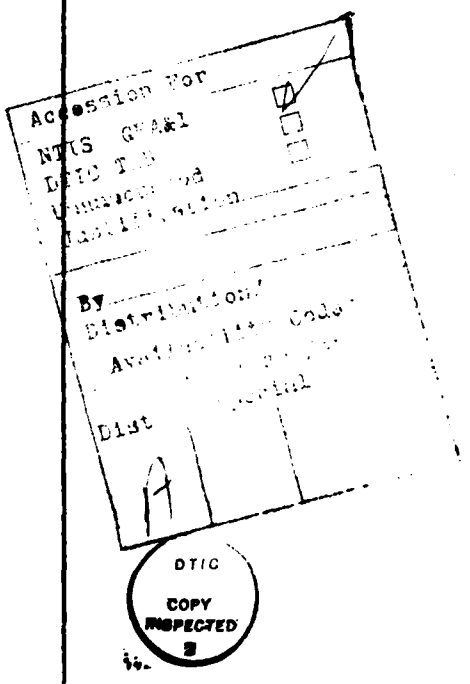
How accurately can the bases, tops, and amounts of the cloud layers over a tactical target area be determined from measurements made with an airborne sensor capable only of detecting cloud presence in situ? Various sampling patterns are evaluated in terms of accuracy of cloud inference as well as cost in fuel and time. Surprisingly, it is found that the best strategy is to sample the volume in alternate ascents and descents in tight spirals, thereby producing a regular pattern of widely separated point measurements in each hori-

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

zontal plane over the target area. The inherent redundancy of closely spaced samples of cloudiness results in a "wastage" of fuel in horizonal flight that more than compensates for the higher rate of fuel consumed in multiple ascents.



PREFACE

Besides the authors of this report, other collaborators in the study were G. J. Higgins, I. A. Lund, P. G. Shapiro, and R. E. Hood. R. F. Fournier of the Air Force Geophysics Laboratory (AFGL) and G. A. Dengel wrote the program required to obtain satellite imagery data, and J. M. Powers collected the data. The report was illustrated by R. Sizer and typed by D. M. Connor. R. F. Wachtmann contributed a painstaking and constructive review of the manuscript.

Special thanks are due to I. I. Gringorten (AFGL) for many illuminating discussions and, particularly, for allowing and facilitating the use of his unpublished model for the vertical distribution of cloud cover.

Throughout the study, helpful guidance was provided by A. M. Gerlach, Program Manager, F. J. Brousaides, AFGL Contract Manager, and M. J. Kraus, AFGL Program Manager.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. OVERVIEW

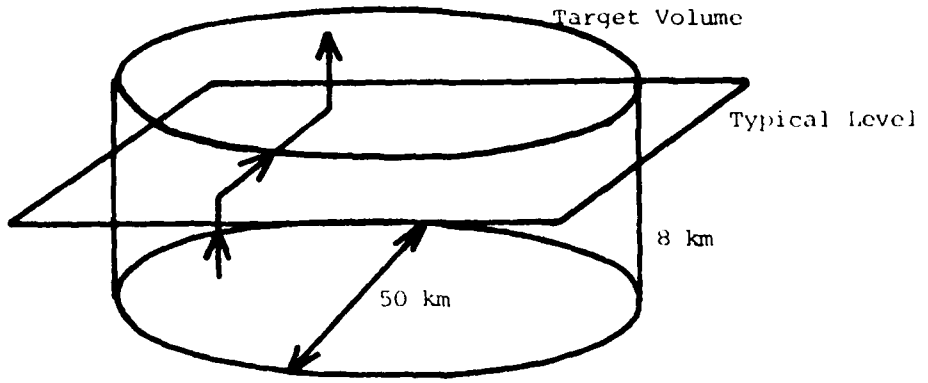
The study reported here was conducted in response to a contract that set the following objective: "Utilizing available climatological cloud property data bases from a representative selection of climates, determine the information content in a series of simulated upper air in situ measurements of cloud presence along a path to provide for estimates of mesoscale cloud cover, tops, and bases from the earth's surface to 8 km above ground level."

The general context of the problem is a battlefield. The specific scenario is a tactical target area 50 km across. The cloud probe is a simple binary sensor, capable of reporting only "I am in cloud" or "I am in clear air." Its reports are telemetered back to a ground station at the rate of 1 Hz. The sensor is carried aboard an automatically piloted vehicle (APV) that is limited in range and speed but is capable of executing a prescribed flight path. Its position is known at all times.

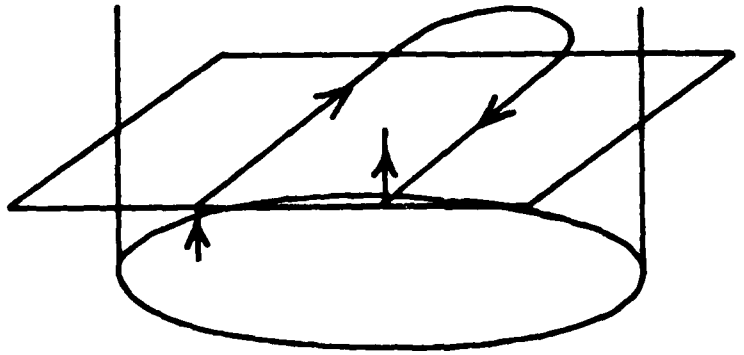
Figure 1.1 illustrates the three general classes of sampling patterns that were examined. In the first, the target volume is sampled through a succession of horizontal patterns that are stepped in altitude, each horizontal sample consisting of measurements taken along a single pass. The second pattern is identical, except that the horizontal sample is now taken along a flight path that is more than a single pass. The third pattern consists of alternate ascents and descents in a tight spiral, the result being that each level is sampled in a pattern of widely separated points.

To evaluate the "sampling accuracy" of these patterns, i.e., the accuracy of inferences based on the samples, we adopted two independent approaches. The experimental approach was founded on a set of 132 actual cloud fields as observed from a Geostationary Operational Environmental Satellite (GOES). Computer programs enabled simulated sampling patterns to be flown through these fields. Analysis of these led to estimates of the sampling accuracy of the various patterns. Details of the basic cloud fields and their collection are described in Appendix A.

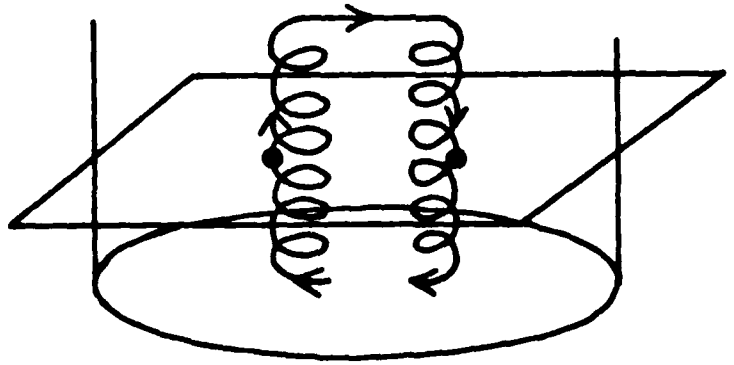
In parallel, theoretical estimates of sampling accuracy were derived from the binomial distribution. This distribution is known to depict accurately the statistical properties of a collective of binary samples, which is what a set of our yes/no cloud measurements comprises. The sole



(a) Horizontal Pass



(b) Horizontal Pattern



(c) Vertical Pattern

Figure 1.1. The Three Classes of Sampling Strategy.

complication is that the binomial distribution is valid only for collectives of independent samples, whereas the successive and nearby point measurements in our scenario are assuredly not independent. In order to retain use of the binomial distribution, we introduce the concept of an "independence fraction," which reduces the N actual points of a sample to a statistically equivalent sample consisting of N' independent points.

Appendix B elaborates on the use of the binomial distribution for our purpose, while Appendix C discusses evaluation of the independence fraction.

Besides the broad objective cited above, the contract Statement of Work (SOW) asks that several pointed questions be answered for the case that the sampling runs are straight line horizontal paths. Then, the Statement asks whether the best estimates of cloud parameters are achieved through this sampling mode or whether some alternate trajectory could significantly reduce uncertainties in these estimates.

In Section 2 of this report, the accuracy of sampling in horizontal passes is evaluated, and its dependence on the length of the pass established. There also, the specific questions of the SOW are addressed one by one.

Section 3 then examines the accuracy of sampling in horizontal patterns other than straight passes and finds that, for a given sampling length, certain patterns are more effective than a straight pass.

Section 4 treats the last of the 3 classes of sampling strategy — the "vertical" flight pattern that produces a configuration of isolated point samples at each of the horizontal levels. Here it is found that, in terms of the accuracy with which the areal cloud fraction can be inferred, a pattern of relatively few, well-positioned points is the equivalent of a rather long sampling path. This is consistent with the estimate of independence fraction of a "continuous" horizontal sample, which suggests that the information content of a 10-km segment is no more than that of its two end points.

Finally, in Section 5 the best of horizontal and vertical sampling patterns are competed in terms of time and fuel required for execution. It is found that, for a given accuracy, the best horizontal strategy costs almost twice as much as the best vertical strategy. Clearly, the waste of time and fuel resulting from the redundancy of information in

horizontal sampling more than compensates for the higher rate of fuel consumption entailed in the "pogo-stick" flight path of the vertical pattern.

Besides the aforementioned advantage of the vertical strategy, it is vastly superior to the horizontal with respect to fixing cloud base and top and, consequently, in recognizing the existence of discrete layers. This latter capability makes possible a confident answer to an important question that can only be guessed at from horizontal samples — namely, what is the overall cloud fraction when more than a single layer is present?

The contract SOW invited us, first, to work the overall problem assuming a horizontally homogeneous cloud field and, then, to consider and evaluate the effects of inhomogeneity. However, neither our experimental approach nor the theoretical was made simpler by an assumption of homogeneity. Consequently, the general case was attacked from the outset, and the conclusions are valid without regard to degree of homogeneity. Nevertheless, Appendix D touches on the academic issue of sampling a homogeneous field and, additionally, discusses a realizable situation that represents, in our view, the most troublesome form of inhomogeneity.

Sampling efficiency is found to depend on the climatological frequency of cloud amount. What is relevant is cloudiness at the level being sampled, not total cloudiness. It is the latter, unfortunately, that is treated in standard climatological summaries. To generate the specialized statistics required for our purpose, a model developed recently by I. I. Gringorten of the Air Force Geophysics Laboratory (AFGL) was employed.

Throughout the study, whether the sampling pattern is horizontal or vertical, what is sampled is the "projected cloud fraction" or "earth cover," in distinction to "sky cover," which is the fractional coverage as seen from a point on the ground. However the SOW poses its questions in terms of cloud cover. Hence, a means of converting from cloud fraction to cloud cover is required and is dealt with in Section 6. There it is found that the differences detectable between the two measures are small relative to the scatter in our data, and we conclude that in operational practice it is better to assume that sky cover is identical to earth cover.

As is commonly the case, some of the intermediate results achieved during the course of this study were not directly used in the final results. Nevertheless, a few of these incidental results are possibly of general interest and are, therefore, described in the two final appendicies: E and F.

2. SAMPLING IN HORIZONTAL PASSES

2.1 Accuracy of Cloud Amount Estimate vs. Length of Sampling Path.

The first of the specific problems posed by the contract Statement of Work (SOW) was: "Assume a straight line vehicle trajectory at a given level through a cloud deck and determine the trade-offs between path length and uncertainty estimates in the calculation of cloud cover." We approached this question both experimentally and theoretically.

2.1.1 The Experimental Answer.

The experimental basis of the entire study is 132 cloud fields observed from a NOAA GOES satellite (National Oceanic and Atmospheric Administration, Geostationary Operational Environmental Satellite). Details of this data base are described in Appendix A. Out of the 132 cases, 50 were randomly selected and set aside as independent data to be used in testing any conclusions based on the "development sample" of 82 cases.

Each of the basic cloud fields is a rectangular array of binary pixels — i.e., picture elements denoting only cloud or no-cloud — covering an area 100x100 km on the earth's surface. The number of rows and columns in the array varies with distance from the sub-satellite point, but the average spacing in our development sample is 1.23 km in the N-S direction and 0.82 km in the E-W.

To answer this first question of the SOW, passes of various lengths from 10 to 100 km were simulated in the observed cloud fields. The procedure will be illustrated for the case of 60-km passes in a cloud field centered over south central Tennessee on December 29, 1980 and observed at local noon. On this occasion the 100x100 km cloud array consisted of 84 rows and 125 columns, and the cloud fraction over the entire array, denoted NA, was 0.444. Along each of the 209 lines of the array (rows and columns together) 3 60-km passes were laid out symmetrically. The cloud fraction, NL, for each of these 627 simulated passes was evaluated, and a frequency distribution constructed. The result is shown in Table 2.1. From this distribution an accuracy index, denoted P(.1), was evaluated. P(.1) is defined as the fraction of the 627 values of NL falling within 0.1 of 0.4 (NA rounded to nearest tenth). In the present case P(.1) = .530.

TABLE 2.1 FREQUENCY DISTRIBUTION OF PASS CLOUD FRACTION, NL.

NL (tenth)	0	1	2	3	4	5	6	7	8	9	10
Frequency (%)	1.2	2.8	7.3	11.0	18.9	23.1	17.5	11.1	5.1	1.5	0

The development sample of cloud fields yielded 82 such values of P(.1). These were grouped according to value of NA and then averaged. The sample standard deviation was also evaluated for each class containing more than 6 values of P(.1).

All told, this procedure was used to generate statistics for simulated sampling passes of 6 lengths: 10, 20, 40, 60, 80, and 100 km. The results are shown in Table 2.2. The strong dependence of P(.1) on NA is strikingly evident in Figure 2.1.

TABLE 2.2 MEAN SAMPLING ACCURACY, P(.1), AND STANDARD DEVIATION, σ, AS A FUNCTION OF PASS LENGTH AND AREAL CLOUD FRACTION, NA.

	10 km		20 km		40 km		60 km		80 km		100 km	
NA (tenths)	P(.1)	σ	P(.1)	σ	P(.1)	σ	P(.1)	σ	P(.1)	σ	P(.1)	σ
0	.96	-	.97	-	.96	-	.98	-	.99	-	.99	-
1	.88	.03	.88	.04	.90	.04	.92	.03	.94	.03	.95	.04
2	.24	.15	.35	.13	.47	.10	.54	.09	.61	.10	.70	.11
3	.21	.08	.31	.10	.43	.14	.54	.14	.58	.16	.64	.15
4	.17	.08	.25	.08	.35	.11	.45	.14	.53	.17	.61	.17
5	.17	.09	.26	.10	.35	.14	.42	.14	.51	.14	.56	.13
6	.17	.10	.24	.13	.33	.14	.42	.14	.51	.14	.60	.14
7	.20	.05	.28	.09	.41	.12	.51	.15	.59	.15	.65	.19
8	.20	.06	.33	.09	.50	.08	.60	.11	.67	.12	.74	.12
9	.85	-	.83	-	.83	-	.84	-	.84	-	.82	-
10	.99	-	.99	-	1.00	-	1.00	-	1.00	-	1.00	-
Unweighted Mean	.46		.52		.59		.66		.71		.75	

In view of this dependence, simple averaging of P(.1) across the values of areal fraction will not produce the correct value of sampling accuracy that can be expected on average when the particular sampling

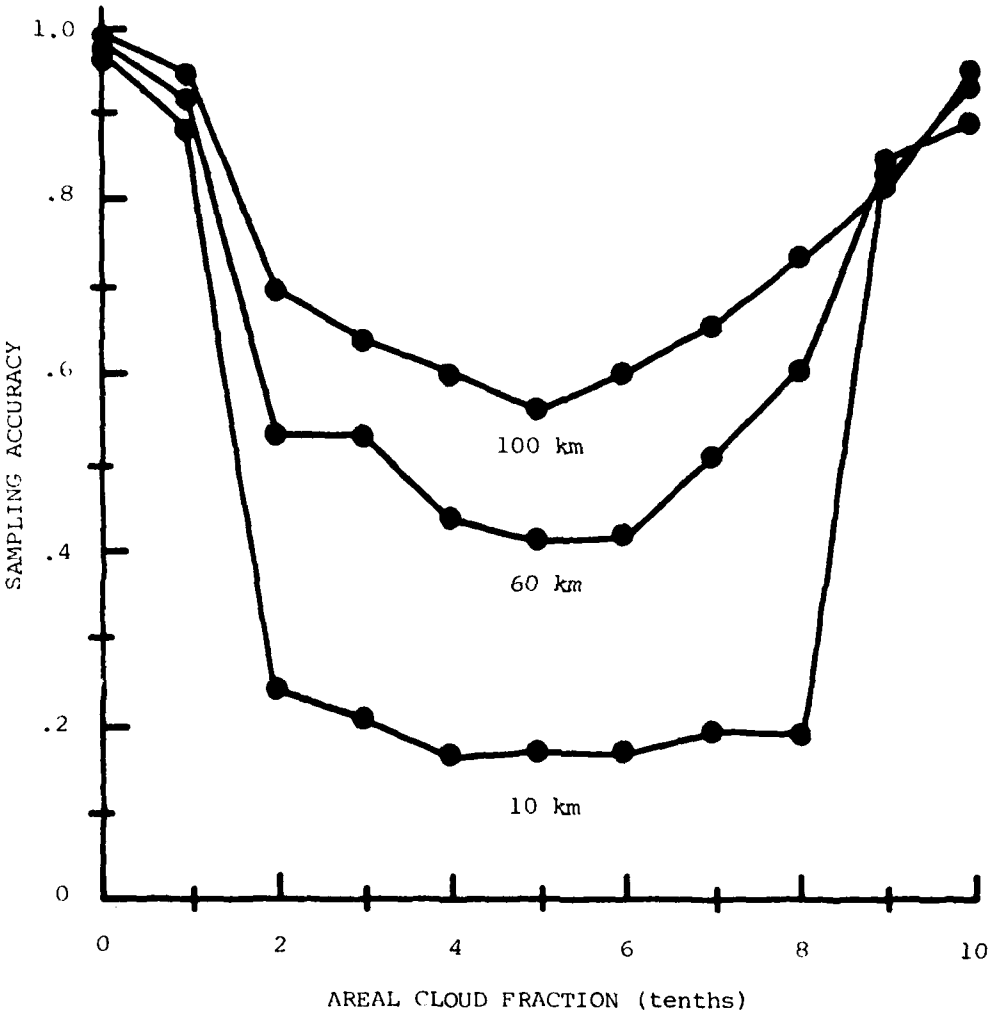


Figure 2.1. Sampling Accuracy, P(.1), as a Function of Areal Cloud Fraction for Three Pass Lengths.

mode is applied on a day-to-day basis or randomly in time. Instead, the averaging must be weighted by the climatological frequency of areal cloud fraction at the level being sampled, which, of course, varies with location and season. If the climatological frequency happens also to be U-shaped, which is a common situation, the expected sampling accuracy can be dramatically better than an unweighted average. This is illustrated in Figure 2.2 which also depicts the trade-off between sampling pass length and accuracy.

The climatological frequencies used in Figure 2.2 are included in Table B-4 of Appendix B. The frequencies for Fulda are for cloudiness in the altitude range 0-3,000 feet, for January, 1200-1400 hours local. These data were derived by Lund using a technique devised by Gringorten.[1] Strictly speaking, the Ft. Rucker data used in Figure 2.2 are not appropriate since they relate to total sky cover, not to cloudiness at a particular level. They were used, nevertheless, in order to demonstrate the effect of a climatology that is not so strongly U-shaped.

We shall be using P(.1) throughout this report, but other investigators have employed the standard error of estimate as their figure of merit for sampling. No simple conversion exists between the two measures, but in Section 4 a relationship between P(.1) and the standard error of regression will be shown.

Figure 2.2 embodies the desired trade-off between path length and uncertainty in the estimate of cloud fraction, but several underlying features warrant emphasis:

A. The simulated passes were located in almost all possible positions within the 100x100 km cloud field. Consequently the values of $\overline{P(.1)}$ represent the expected accuracy of a pass that is <u>randomly</u> positioned in the target area. In the next section we consider patterns that are deliberately positioned relative to the target and find some that are more accurate than random passes of the same length.

B. In order to accommodate the longer passes, it was necessary to deal here with the entire 100x100 km cloud field, rather than a quadrant, which is the size specified for the target area.

---

1. Gringorten, I. I., 1981: Climatic probabilities of the vertical distribution of cloud cover. AFGL-TN (in press).

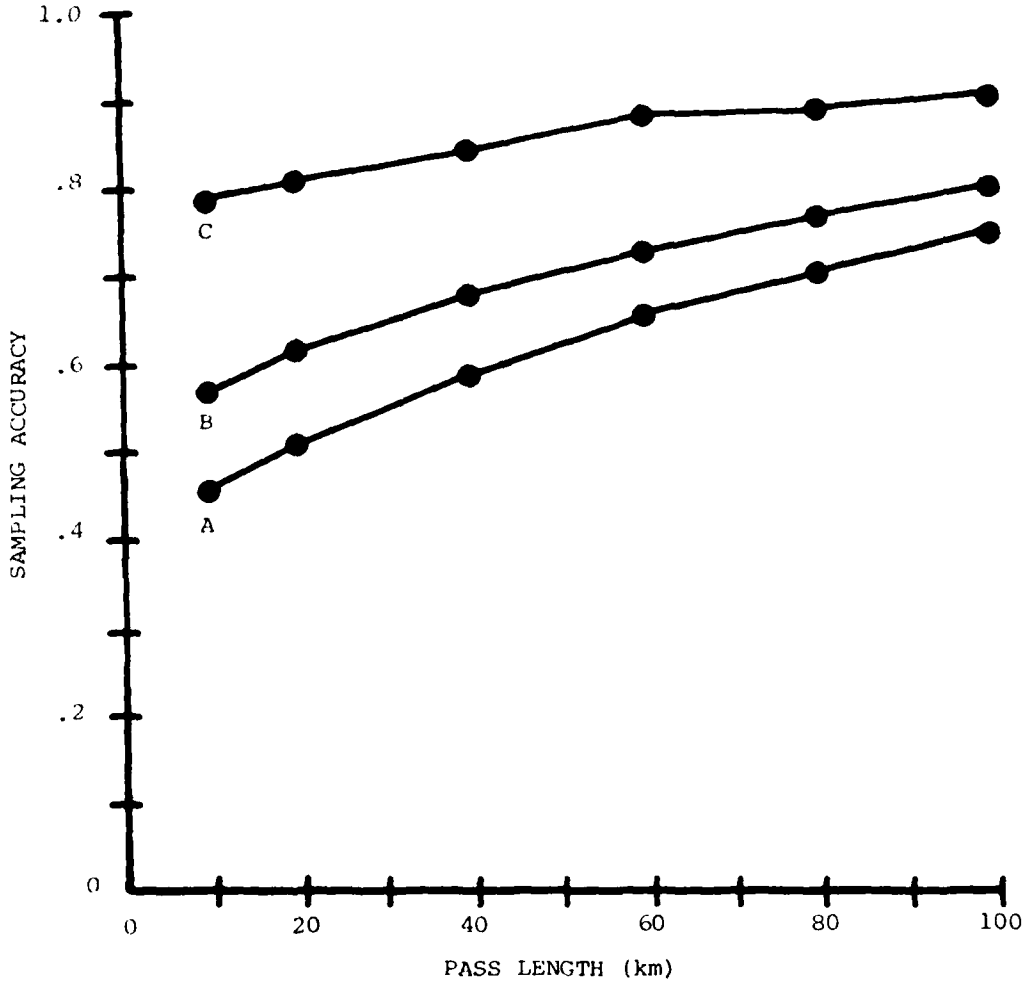


Figure 2.2. Average Sampling Accuracy, $\overline{P(.1)}$, as a Function of Pass Length.

A. Unweighted Average.
B. Average Weighted by Cloud Frequency at Ft. Rucker, AL, in July.
C. Average Weighted by Cloud Frequency at Low Altitude at Fulda, FRG, in January.

2.1.2 The Theoretical Answer.

If the airborne cloud sensor is sampled at the rate of 1 Hz, a horizontal pass yields a set of point samples separated by less than 50 meters. Appendix B outlines how the binomial distribution can be used to determine the accuracy with which areal cloud fraction can be estimated from the cloud fraction observed on a set of points. The sole obstacle to immediate application of this theory in our scenario is that the theory calls for mutually independent samples whereas, due to the spatial coherence of cloudiness, our closely spaced samples are not at all likely to be statistically independent.

By-passing this complication for the moment, let us examine how sampling accuracy, P(.1), depends on sample size. The data plotted in Figure 2.3(A) for samples of 5, 10, 15, and 20 independent points were derived according to the procedure of Appendix B. Just like the experimental values of P(.1), the theoretical values are sensitive to areal cloud fraction. Hence, Figure 2.3(A) displays averages of P(.1) weighted by the same climatologies used in the preceding paragraph. Thus, Figure 2.2 and Figure 2.3(A) are fully analogous and could be directly compared were it not for the *difference* in abscissas: "sample length" in one case, "number of points" in the other.

To rectify this incompatibility we invoke the concept of "independence length" introduced in Appendix C. This is the distance of separation that is sufficient to insure that cloud samples are statistically independent. The average value of this length evaluated on our development sample of data is 12.33 km. This value is used to convert the sample size (number of points) in Figure 2.3(A) into an equivalent length of sampling pass. In Figure 2.3(B) the results of this conversion are plotted, together with the points of Figure 2.2.

Figure 2.3(B) offers an extended view of the trade-off between pass length and sampling accuracy and reveals a compatibility between the experimental and theoretical values of $\overline{P(.1)}$.

2.1.3 Predictive Value of the Sample Autocorrelation.

According to preceding paragraphs, the accuracy of an inferred value of areal cloud fraction depends on the equivalent number of independent points in the linear sample, and this number depends on the "independence

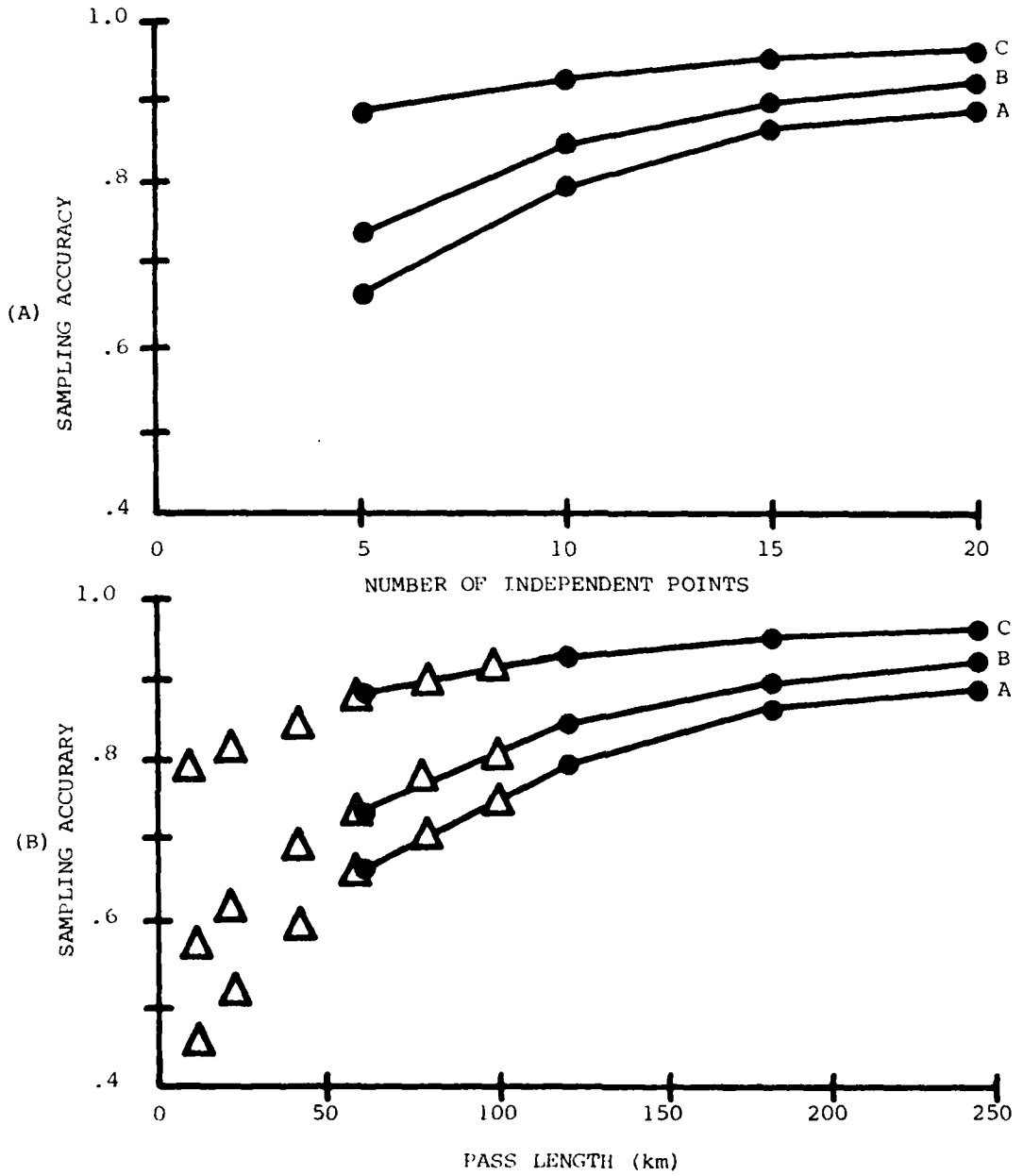


Figure 2.3. Average Sampling Accuracy, $\overline{P(.1)}$, as a Function of Sample Size (A) or Pass Length (B).

A. Unweighted Average.

B. Average Weighted by Cloud Frequency at Ft. Rucker, AL, in July.

C. Average Weighted by Cloud Frequency at Low Altitude at Fulda, FRG, in January.

•——• From Theoretical Model.

Δ From Cloud Data (Figure 2.2).

length." It is reasonable to expect that the independence length and, therefore, the intrinsic sampling accuracy vary in value from day to day. Although the independence length is evaluated empirically in Appendix C, it can be formulated theoretically, and the sample autocorrelation function plays a key role in this formulation.

This line of reasoning led to an experiment to test whether the areal cloud fraction can be more accurately determined if both the 1-lag autocorrelation coefficient of the sampling pass and its cloud fraction are used as predictors.

Each of the 82 observed cloud fields was subdivided into 50x50 km quadrants. For each row and each column of every quadrant, the cloud fraction and the 1-lag autocorrelation, ρ_1, were evaluated — also, the areal cloud fraction for the quadrant. The same was done for the full 100x100 km area. Altogether, the result was more than 34,000 pairs of line and areal cloud fractions, together with values of ρ_1 for the line. These pairs were stratified by value of ρ_1 into 4 classes: less than 0.5, 0.5-0.9, greater than 0.9, and all values. For each class of ρ_1 and for each value of linear cloud fraction, the distribution of areal cloud fraction was determined, along with a variety of statistics.

Figure 2.4 shows how the mean areal cloud fraction, NA, varies with the linear fraction, NL, for the 4 classes of ρ_1. Table 2.3 shows how the average sampling accuracy, P(.1), varies among the classes. In both instances, the variation with class is no more than might be expected as a sampling fluctuation. We conclude, therefore, that the predictability of NA from NL is negligibly enhanced by knowledge of ρ_1.

2.2 Optimum Number of Levels to Sample.

The second problem raised by the Statement of Work was: "Determine the optimum number of levels which the detector should traverse, consistent with restrictions in vehicle range, in order to characterize areal cloud coverage."

The crux of the matter here is to strike the best compromise between accuracy of cloud amount at each level and assurance that no layer goes undetected. The former is best served by maximizing the pass length at each level, the latter by maximizing the number of levels sampled.

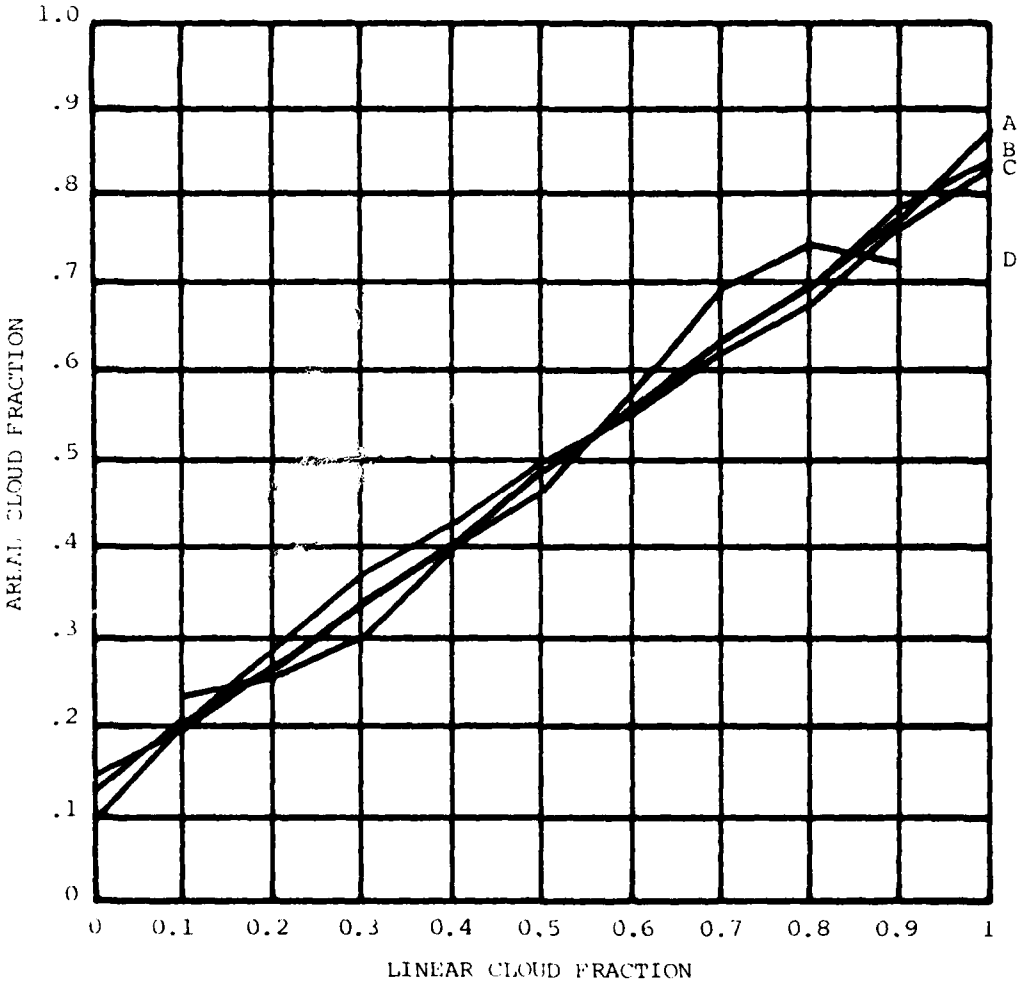


Figure 2.4. Areal Cloud Fraction as a Function of Linear Fraction for Various Classes of the 1-Lag Autocorrelation of the Line Sample.

A. All ρ_1.

B. $.5 \leq \rho_1 < .9$.

C. $\rho_1 < .5$.

D. $\rho_1 \geq .9$.

TABLE 2.3 MEAN SAMPLING ACCURACY, P(.1), AS A FUNCTION OF LINEAR CLOUD FRACTION, NL, FOR VARIOUS CLASSES OF THE 1-LAG AUTOCORRELATION, ρ_1, OF THE LINE SAMPLE.

NL (tenths)	P(.1) $\rho_1 < 0.5$	$0.5 \leq \rho_1 < 0.9$	$\rho_1 \geq 0.9$	All ρ_1
0	.66	.63	*	.76
1	.66	.71	.60	.68
2	.70	.66	.78	.74
3	.65	.72	.78	.71
4	.73	.67	.47	.66
5	.69	.59	.65	.61
6	.73	.60	.45	.61
7	.69	.67	.71	.68
8	.63	.65	.76	.65
9	.57	.66	.54	.61
10	.55	.66	*	.69
Average	.66	.66	.64	.67

In the absence of foreknowledge as to the types of clouds likely in the target area, the best strategy is to fly passes no longer than 10 km and to sample as many levels as possible, uniformly distributed within the altitude range of prime operational interest. The reasons are the following:

A. As shown in Figure 2.2, sampling accuracy improves only slowly with pass length, particularly in climates like that of Fulda.

B. An inference of areal cloud coverage is damaged far more by failure to detect a layer altogether than by a degraded estimate of its amount.

C. The frequency of cloud occurrence is typically a weak and quasi-monotonic function of altitude.

An attempt could be made to quantify this solution, but we chose not to because we doubt that there is a "good" answer for the case of horizontal sampling and, more important, because the problem does not even exist for a distinctly superior sampling mode that is treated in Sections 4 and 5.

2.3 Uncertainty of Tops and Bases.

The third specific question posed by the Statement of Work was: "Determine the uncertainty in the estimation of tops and bases given measurements on a straight line trajectory at multiple flight levels."

In the given circumstances, the uncertainty in estimate of the top/base of a cloud layer is half the distance between sampling levels. A cloud top/base would be declared to exist whenever cloud is detected on one pass but not on the next higher/lower pass. To first approximation, the median position for the boundary is the midpoint between the two sampling levels.

A pedantic refinement of this estimate would take account of the fact that whereas the existence of cloud at the one level is 100% certain, it is less than dead certain that the other level is clear. The degree of uncertainty depends on the length of sampling pass and can be estimated by means described in Section 2.1. This would lead to biasing the median estimate of the boundary altitude toward the level at which no cloud was detected.

Again, it is fortunate that this problem evaporates in the alternative sampling strategy recommended in Section 5.

3. SAMPLING IN HORIZONTAL PATTERNS

The averages of sampling accuracy presented in Section 2.1.1 are based on all possible locations of passes, which include, for the 100x100 km area, passes that are as remote as 50 km from the "target." Consequently, as previously noted, the findings of that section characterize the performance expected of sampling on horizontal passes that are randomly positioned in the target area. It is reasonable to expect that centrally located passes might be more representative of the area. Also, for a fixed allocation of fuel to sample a level, it might be more efficient to spend this on several short passes rather than in a single long pass across the area. With such thoughts in mind, we designed a series of experiments on the observed cloud fields to test whether sampling in a prescribed horizontal pattern is more efficient than the same distance of sampling in a pass that is randomly located in the target area.

The patterns tested are shown in Figure 3.1. The results to be quoted for configurations #1, 2, and 3 combine the row-patterns illustrated here and the analogous column-patterns which are not shown. In all cases the area sampled is 50x50 km. The sampling length is 100 km for all patterns except for #1 whose length is 150 km. Except for the closed pattern of #5, the total flight path would have to be longer than the sampling path in order to link the passes. In configurations #1 and #3 the sampling passes subdivide the area respectively into 4 and 3 equal parts. #2 is an example of "equal-area" sampling, which will also play a key role in Section 4. To construct #2, the area was first divided into halves; then a sampling pass was made through the middle of each half. The square sampling pattern in #5 measures 25x25 km.

For the test, the linear cloud fraction was measured for each of the 8 configurations for all 4 quadrants of each of the 82 100x100 km cloud fields. (There are 8 configurations, rather than 5, because the row and column variants of numbers 1, 2, and 3 were treated separately at this stage.) Thus, for each configuration there were 328 pairs of linear and areal cloud fractions. These were stratified according to linear fraction in tenths, and a frequency count was made for the areal fraction. An example of the resulting matrix is shown in Table 3.1.

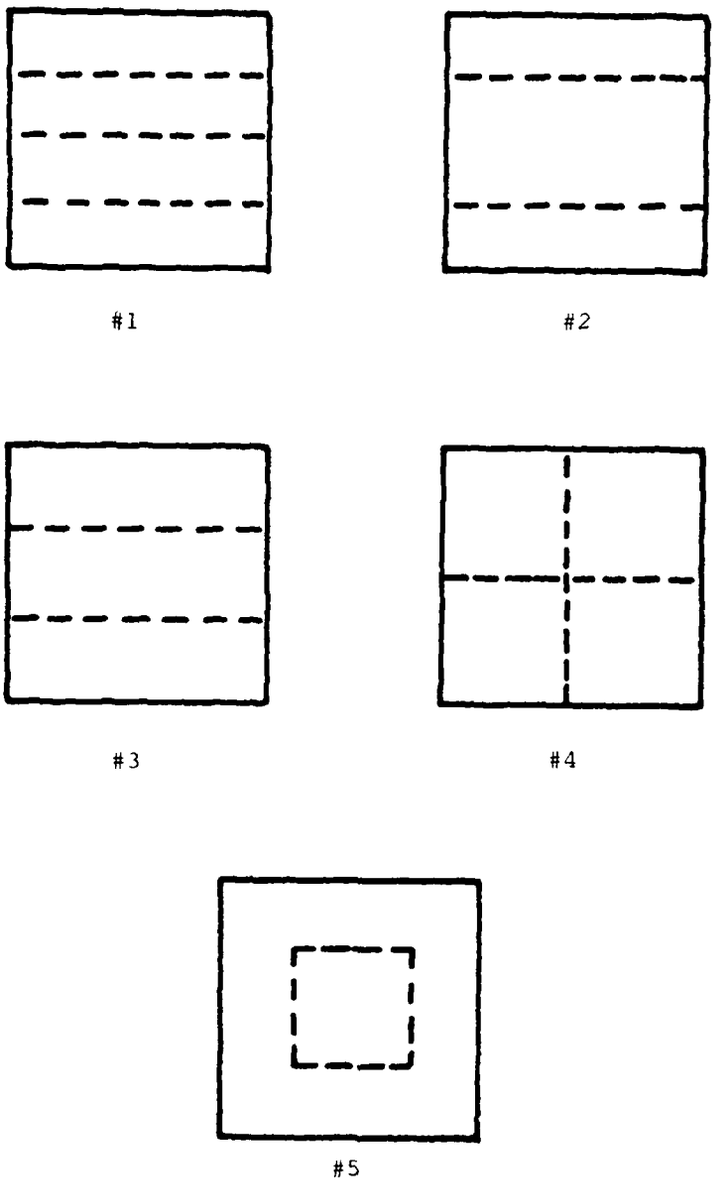


Figure 3.1. Horizontal Sampling Patterns.
(Dashed Lines).

TABLE 3.1 FREQUENCY (%) OF AREAL CLOUD FRACTION, NA, AND SAMPLING ACCURACY, P(.1), AS A FUNCTION OF LINEAR CLOUD FRACTION, NL, FOR CONFIGURATION NO. 5. N IS THE NUMBER OF CASES.

NL \ NA	0	1	2	3	4	5	6	7	8	9	10	N	P(.1)
0	65	28	7	0	0	0	0	0	0	0	0	43	.93
1	5	47	34	11	3	0	0	0	0	0	0	38	.86
2	0	11	40	40	9	0	0	0	0	0	0	35	.91
3	0	3	18	41	38	0	0	0	0	0	0	34	.97
4	0	0	0	17	38	38	7	0	0	0	0	29	.93
5	0	0	0	16	26	32	21	5	0	0	0	19	.79
6	0	0	0	0	12	21	39	24	3	0	0	33	.84
7	0	0	0	0	3	6	29	55	6	0	0	31	.90
8	0	0	0	0	0	0	13	33	46	8	0	24	.87
9	0	0	0	0	0	0	5	5	35	50	5	20	.90
10	0	0	0	0	0	0	0	0	9	50	41	22	.91
											AVERAGE		.892

(Because of rounding to integral values of percent, the frequencies do not always sum to 100.)

From these distributions the sampling accuracy, P(.1), was evaluated for the 11 values of linear fraction, and these were then averaged (without weighting for climatological frequency). Finally, the row and column results for configurations #1, 2, and 3 were averaged.

The final results are displayed in Table 3.2. The best accuracy is achieved by pattern #1, but it is 50% longer than the others. Among the 100 km patterns, #2 is the best. As previously noted, it is the equal-area pattern. In Section 4 it will be seen that equal-area configurations are the most efficient for point sampling also. #5 is less accurate than #2 or #3, but it is more economical of fuel.

TABLE 3.2 AVERAGE SAMPLING ACCURACY, $\overline{P(.1)}$, OF HORIZONTAL PATTERNS.

(Configurations 1, 2, and 3 include passes along columns as well as the row patterns shown in Figure 3.1.)

CONFIGURATION #	LENGTH	$\overline{P(.1)}$
1	150 km	.972
2	100	.960
3	100	.939
4	100	.863
5	100	.892

As speculated in the introduction to this section, all of the patterns here are more effective for sampling than the random passes treated in Section 2. The unweighted average of P(.1) for 100 km in Figure 2.2 is .751, but this is an unfair comparison because the area involved there is 100x100 km, whereas here a 100-km linear sample is used to specify the cloud fraction for a 50x50 km area. A more apt comparison, albeit somewhat artificial, would be with the accuracy of a 200-km random pass for the 100x100 km area. An estimate of this, from the theoretical results shown in Figure 2.3, is .871.

The principal findings of this section are:

A. When horizontal passes are used to specify the cloudiness over a target area, it is more efficient to pattern and position the passes than to sample along paths randomly located in the target area.

B. For a fixed total length of sampling path, an "equal-area" pattern is more effective than the others examined here.

SECTION 4. SAMPLING IN VERTICAL PATTERNS

The comparisons drawn in Section 2.1.2 highlight the fact that horizontal sampling is intrinsically inefficient, owing to the strong autocorrelation of cloudiness. For instance, sampling along 60 km has the predictive power of only 5 independent point samples. The time and fuel spent traversing the distance between points that are 12 km apart are effectively wasted.

This observation inspired a series of experiments to measure the relative efficiency of point samples in regular patterns, such as would result if the target volume were sampled in tight vertical spirals connected by short horizontal segments at top or bottom, as illustrated in Figure 1.1(C).

The 15 configurations tested are shown in Figure 4.1. For all, the area is a 50x50 km quadrant of the basic cloud field; hence there are 4 x 82 = 328 cases for each configuration. Corresponding to configurations #1-5 there are half-scale counterparts (#6-10) which are not illustrated. In these, the identical pattern is arrayed over the co-centered square measuring 25x25 km. In all instances, including the half-scale configurations, it is the cloud fraction of the 50x50 km area that we are trying to specify from the cloud fraction of the point sample.

Patterns #12, 13 and 14 are "equal-area" in the sense introduced in the preceding section. Each sample point is the center of one of the N equal squares into which the quadrant is subdivided. As we shall be seeing, these are the most efficient of the configurations tested.

For each configuration, there are 328 pairs of point-sample cloud fraction, NP, and areal cloud fraction, NA. These were stratified according to the point fraction in tenths, and a frequency distribution was constructed for the areal fraction. A typical result, that for configuration #13, is shown in Table 4.1. From these distributions the sampling accuracy, P(.1), was evaluated for the 11 values of point fraction, and these were then averaged without weighting.

In addition, for each configuration NA was linearly regressed on NP, using the 328 pairs of values. Subsequently the regression equations were tested on the independent sample of 4 x 50 = 200 cases.

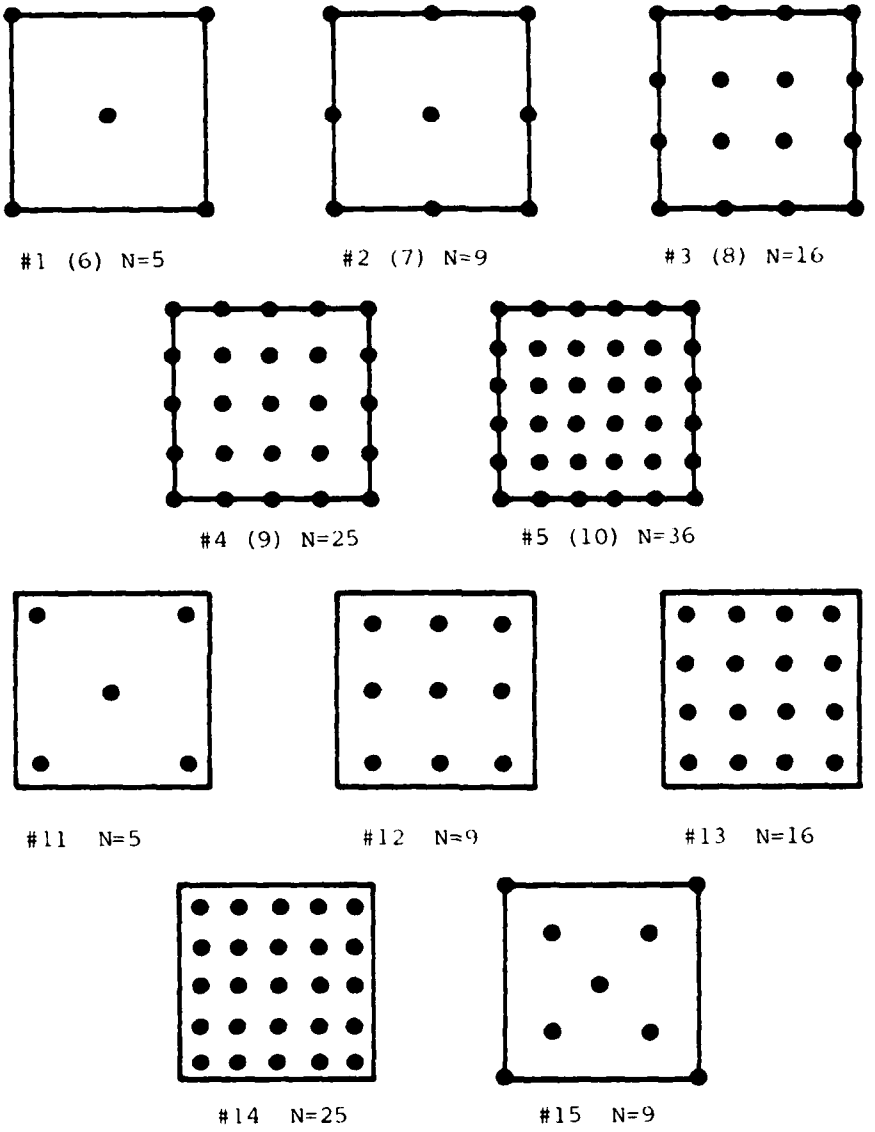


Figure 4.1. Point-Sampling Configurations. #1 (6) Means That Configuration #6 Is the Half-Scale Version of #1. See Text for Further Details.

TABLE 4.1 FREQUENCY DISTRIBUTION (%) OF AREAL CLOUD FRACTION, NA, AS A FUNCTION OF POINT CLOUD FRACTION, NP, FOR CONFIGURATION NO. 13.

NP and NA are, respectively, point and area fractions (tenths); N is number of cases; P(.1) is sampling accuracy (%).

NP \ NA	0	1	2	3	4	5	6	7	8	9	10	N	P(.1)
0	71	29	0	0	0	0	0	0	0	0	0	28	100
1	23	55	20	2	0	0	0	0	0	0	0	44	98
2	0	20	47	27	7	0	0	0	0	0	0	15	94
3	0	0	37	47	40	6	0	0	0	0	0	51	94
4	0	0	2	21	48	19	10	0	0	0	0	42	88
5	0	0	0	10	35	25	25	5	0	0	0	20	85
6	0	0	0	0	13	23	35	28	3	0	0	40	86
7	0	0	0	0	0	5	27	41	27	0	0	22	95
8	0	0	0	0	0	0	7	43	33	17	0	30	93
9	0	0	0	0	0	0	4	4	21	57	14	28	92
10	0	0	0	0	0	0	0	0	0	25	75	8	100

The results of these analyses are plotted in Figure 4.2 as a function of the number of points in the sampling pattern. The sampling accuracy, P(.1), is shown in the upper half of the figure, with the points identified by configuration number. The 3 equal-area configurations, which are connected by lines, are consistently the best in their class.

In the lower half of Figure 4.2 the standard error of regression is plotted. Again the 3 cases of equal-area patterns are superior. The errors shown here are based on the dependent data, but as can be seen in Table 4.2, there is insignificant difference between these and the standard errors based on independent data. In fact, the two sets of errors have identical averages, 1.0 tenths.

The data plotted in Figure 4.2, plus the correlation coefficients, are tabulated in Table 4.2.

Results thus far in this section are based on the realistic assumption that NP is given and NA is to be specified. The 328 pairs of values for each configuration were also analyzed assuming NA given and NP to be predicted. This was done to facilitate a direct comparison, in terms of sampling accuracy, between these configurations and the same number of random points. The results are given in Table 4.3, where the columns

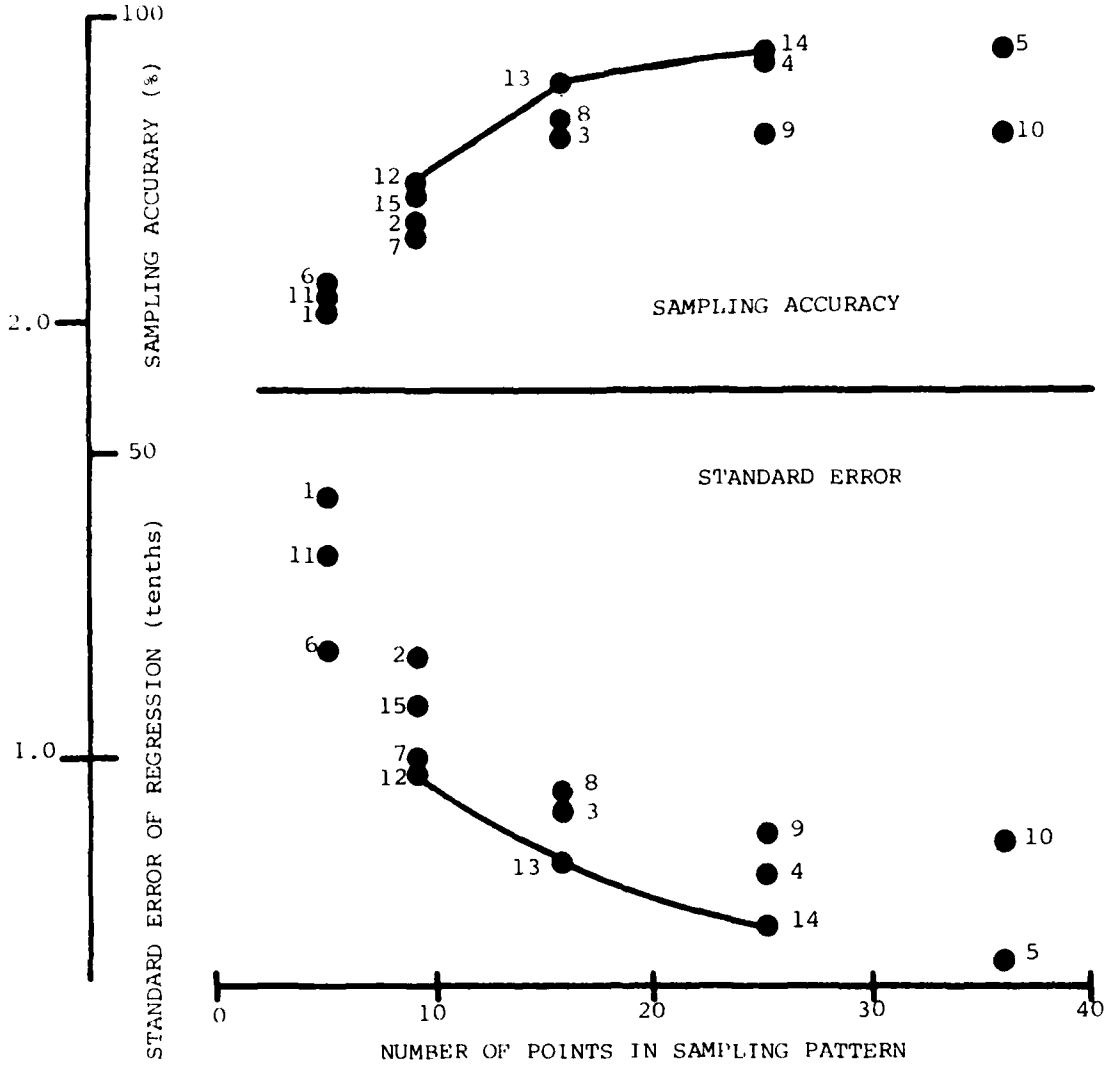


Figure 4.2. Sampling Accuracy and Standard Error of Regression for Point-Sampling Configurations. Lines Link the Equal-Area Patterns.

TABLE 4.2 SAMPLING ACCURACY OF THE POINT CONFIGURATIONS; CORRELATION COEFFICIENT AND STANDARD ERROR FOR LINEAR REGRESSION OF AREAL CLOUD FRACTION ON POINT-SAMPLE CLOUD FRACTION.

Errors I and II are based, respectively, on the dependent and the independent samples of data.

# OF POINTS	5			9				16			25			36	
CONFIGURATION #	1	6	11	2	7	12	15	3	8	13	4	9	14	5	10
P(.1) (%)	67	70	68	77	76	81	80	87	89	93	95	87	96	98	88
CORRELATION	.82	.90	.85	.90	.94	.94	.92	.95	.95	.96	.97	.96	.98	.98	.96
STANDARD ERROR I (tenths)	1.6	1.3	1.5	1.2	1.0	1.0	1.1	0.9	0.9	0.8	0.7	0.8	0.6	0.6	0.8
STANDARD ERROR II (tenths)	1.5	1.4	1.3	1.2	1.1	1.1	1.1	0.9	0.9	0.8	0.7	0.9	0.6	0.6	0.8

TABLE 4.3 SAMPLING ACCURACY OF THE BINOMIAL DISTRIBUTION AND THE POINT CONFIGURATIONS.

# OF POINTS	5				9					16				25				36		
CONFIGURATION #	B	1	6	11	B	2	7	12	15	B	3	8	13	B	4	9	14	B	5	10
P(.1) (%)	76	76	82	78	76	77	79	83	80	82	85	79	90	86	88	78	90	91	93	79

labeled B denote the unweighted average accuracy of random patterns as derived from the binomial distribution. With only a few exceptions, which are mainly for the half-scale patterns, the random patterns are less efficient than all of the systematic patterns. This suggests that none of the full scale patterns suffers from loss of efficiency due to spatial autocorrelation. That circumstance was to be expected for all of the full scale patterns except, possibly, #'s 5 and 14, where the points are separated by only 10 km. In Appendix C, the average value of the "independence length" is found to be 12 km.

The values of P(.1) are different between Tables 4.2 and 4.3 because of the difference in the underlying conditional frequencies: NP being the predictor in 4.2, NA in 4.3. NP is seen to be the more efficient predictor for the larger samples, NA for the smaller.

As mentioned in Section 2, the standard error of estimate is sometimes used as the index of sampling accuracy. Table 4.2 lists both P(.1) and standard error of regression for the 15 configurations. Figure 4.3 reveals that the two measures are well related here, but there is no assurance that this relation is applicable to other sampling patterns.

The principal findings of this section are:

A. Point samples taken in well-distributed patterns are generally more efficient than random point samples. Consequently, the binomial distribution can be used as a conservative estimator of the expected accuracy of systematic patterns of point samples.

B. Equal-area point patterns are the most efficient of the configurations tested.

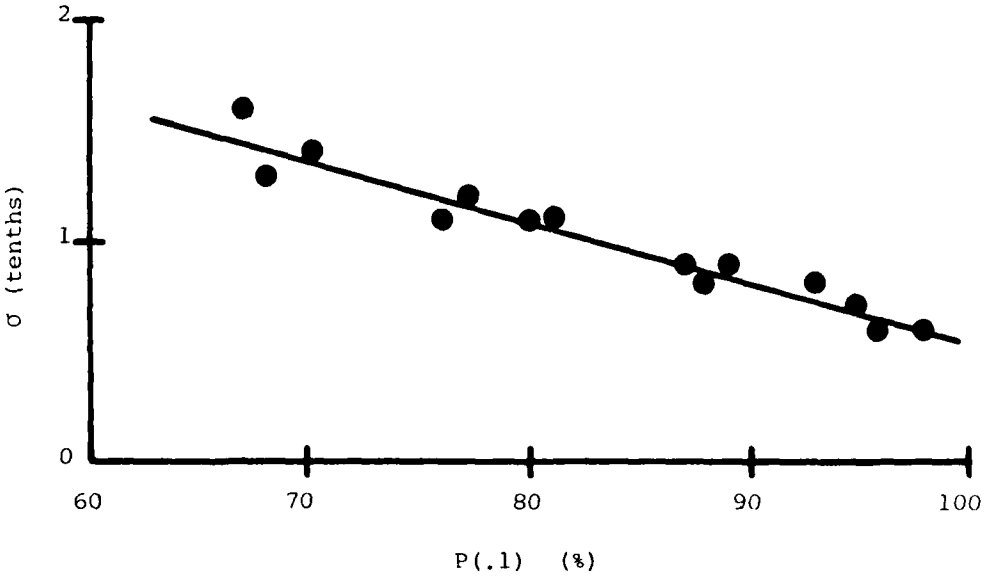


Figure 4.3. Standard Error as a Function of P(.1).

5. LOGISTICAL COMPARISONS AND RECOMMENDATIONS

5.1 Comparison of Horizontal and Vertical Sampling Patterns.

In the two preceding sections it has been shown that the so-called equal-area configurations, whether lines or points, are the most efficient samplers. It remains to consider the logistical efficiency of lines and points; that is, to compare the fuel and time required to execute the horizontal and vertical sampling patterns that are the means of achieving the line and point configurations.

The performance specifications assumed for the sampling platform are shown in Table 5.1. These were provided by Dr. Gerald Seemann, president of Developmental Sciences, Inc., in response to a request for generalized values that are compatible with capabilities of state-of-the-art automatically piloted vehicles (APV's).

The volume to be sampled is 50 km across and from 1,000 to 10,000 feet in the vertical. The horizontal samples are taken at intervals of 1,000 feet, from top to bottom, for a total of 10 levels.

The horizontal patterns chosen for the "flyoff" are #'s 2 and 5 in Figure 3.1, hereafter referred to as H-2 and H-5. Both yield a 100 km sample at each level. H-2 is the efficient equal-area pattern, but it has the drawback of requiring an extra leg of 25 km to close the pattern. For present purposes we ignore any additional information that might be gleaned from the bridging leg. H-5 is intrinsically less accurate than H-2 but requires no extra leg.

The vertical patterns are #'s 12 and 13 in Figure 4.1 — henceforth, V-12 and V-13. They are, respectively, the equal-area 9-point and 16-point configurations. As illustrated in Figure 1.1, they are generated by flying alternate ascents and descents in tight spirals. A standard-rate turn produces a radius of about 0.65 km for the spiral, which is treated here as a vertical line. The spirals are linked at top and bottom by horizontal legs: 17 km long for V-12, 13 km for V-13. Again, we ignore the possibility of sampling on these legs.

The results of the flyoff are posted in Table 5.2, where the patterns are listed in order of decreasing time and fuel consumption.

Because the horizontal patterns are flown from top to bottom, they entail no ascents. The values of sampling accuracy are derived from Tables 3.2 and 4.2, and are cited here only for intercomparison.

TABLE 5.1 PERFORMANCE POSTULATED FOR APV.

FLIGHT MODE	AIRSPEED (KTS)	VERTICAL RATE (FT/MIN)	FUEL RATE (LBS/HR)
HORIZONTAL	70	——	8
ASCENT	65	500	10
DESCENT	70	2000	5

TABLE 5.2 TIME AND FUEL REQUIRED FOR VARIOUS SAMPLING STRATEGIES.

	H-2		H-5		V-13		V-12	
PATTERN SEGMENTS	TIME (HRS)	FUEL (LBS)	TIME (HRS)	FUEL (LBS)	TIME (HRS)	FUEL (LBS)	TIME (HRS)	FUEL (LBS)
HORIZONTAL	9.64	77.09	7.71	61.67	1.45	11.56	1.03	8.22
ASCENT	——	——	——	——	2.40	24.00	1.20	12.00
DESCENT	0.08	0.38	0.08	0.38	0.60	3.00	0.38	1.88
TOTAL	9.72	77.47	7.79	62.05	4.45	38.56	2.61	22.10
P(.1)(%)	96		89		93		81	

As a class, the vertical patterns are the indisputable winners of the flyoff. H-2 does afford the best sampling accuracy, but it is only slightly more accurate than V-13 while it consumes more than twice the fuel and time. Even H-5, which is less accurate than V-13, takes three-quarters more time and uses 60% more fuel. The most economical of the four patterns, V-12, requires less than 60% of the time or fuel consumed by V-13, but at a sizeable penalty in accuracy. However, as will be seen below, even V-12 can be expected to achieve 90% accuracy on average in the real world.

5.2 Refined Estimates of Sampling Accuracy

There are two reasons why the estimates of sampling accuracy in Table 5.2 are pessimistic: A) they are unweighted averages, and B) they

are for cloud levels rather than layers.

Why an unweighted average of P(.1) is an improper and, usually, pessimistic estimator of sampling accuracy is discussed in Section 2 and Appendix B. The correct estimator of the sampling accuracy expectable in routine practice is a climatically weighted average. Figure B-1 suggests that this estimator is strongly correlated with the degree to which the frequency of cloud amount is U-shaped.

Climatically weighted averages of sampling accuracy for the 4 patterns in the flyoff are listed in Table 5.3. Again, Ft. Rucker July and Fulda 1-3 KFT January are the two cloud climatologies used for the illustration. Now, even the "cheap" pattern, V-12, averages 90% in accuracy.

TABLE 5.3 AVERAGE SAMPLING ACCURACY.

	$\overline{P(.1)}$ (%)			
		WEIGHTED		
SAMPLING PATTERN	UNWEIGHTED	FT. RUCKER	FULDA	AVERAGE
H-2	96	98	99	98
H-5	89	93	96	95
V-13	93	95	98	97
V-12	81	88	94	91

These values of accuracy are valid for sampling a level. In practice, a single cloud layer will be sampled at more than one level. This is particularly true for vertical patterns, in which the levels are as close as 33 feet if the sensor is sampled at the rate of 1 Hz. The estimator for the cloud fraction of the layer is the average of the cloud fractions for all levels sampled within the layer. Consequently, the layer cloud fraction is normally more accurate than the level fraction. Just how much more accurate is a question that could be readily answered if the level samples were independent, but they are not, owing again to the spatial coherence of cloudiness. It is beyond the scope of our data to appraise quantitatively the improvement due to averaging of interdependent levels, but it can be taken with confidence that the values in Table 5.3 are conservative estimators for cloud layers of substantial thickness.

5.3 Other Advantages of Vertical Sampling Patterns.

Besides the distinct superiority in fuel economy, vertical sampling patterns have two other significant advantages over horizontal: greater accuracy both in estimates of cloud base/top and in total cloud amount.

In the present scenario, if the cloud sensor is sampled at 1 Hz, cloud heights are determined from the vertical patterns to within 10 feet on ascent, whereas there is an uncertainty of ± 500 feet in the estimates of cloud base or top derived from horizontal sampling.

From vertical samples, the total cloud amount (within the volume sampled) is readily estimated, as simply the fraction of profiles on which cloudiness is detected at any height. This estimate of total cloudiness, as argued above with respect to a layer, is at least as accurate as the estimate of cloud amount at a level, and probably more so.

By contrast, estimating the total cloud amount from horizontal samples depends on answering two sometimes difficult questions: how many cloud layers are present, and how do they overlap? For sampling at 1,000-ft intervals it is reasonable, but not necessarily correct, to assume that any cloudiness encountered on adjacent levels is part of a single layer. Multiple layers are declared to exist only when one or more intervening cloud-free levels are observed. As to the overlap of multiple layers, one has little choice but to assume the cloud elements in separate layers to be independent in their positioning. Clearly, both steps in the process are potential sources of significant error. For example, if 5/10 cloudiness is detected at both of two levels, the estimate of total cloudiness will be either 5/10 or 8/10 depending on whether the two levels are assumed to be part of a single layer or two layers are assumed to be present.

5.4 Recommended Operational Procedure.

5.4.1 Choice of Sample Size.

The unequivocal finding of this section is that a vertical, equal-area sampling pattern is superior. In practice, then, the only choice open is how many points (profiles) there will be in the pattern. This

must be decided as a compromise among accuracy desired, volume within which the cloud parameters are to be determined, and time/fuel available for sampling.

Save for the possibility of using a faster APV, there are only two means of reducing the time required to sample the cloudiness over a target area: A) reduce the altitude range sampled; B) reduce the number of profiles (points) sampled.

Sampling time does not decrease linearly with reduction in altitude range because, even in vertical sampling, the horizontal legs account for a non-trivial fraction of the overall time. In Table 5.2 the horizontal time is at least 1/3 of the total for V-12 and V-13.

Nor does the overall sampling time scale linearly with the number of points in the pattern. While the time required for the vertical components is proportional to the number of points, the time spent on the horizontal component scales almost like the square root of the number of points. The reason for this is that, as the number of points decreases in an equal-area pattern, the separation of the points increases.

The penalty for reducing the number of profiles sampled is, of course, reduced accuracy of the estimates of cloud amount. If the sampling area is 50 km across, then the values of P(.1) cited in Table 5.3 are valid estimates for the areal cloud fraction at a level. However, as noted above, these are conservative estimates for the accuracy of the cloud fraction of a layer or for the accuracy of total cloud amount.

Accuracies achievable with samples sizes other than 9 or 16 points can be derived by interpolation/extrapolation from Table 5.3. For this purpose Figure 2.3(A) is a useful guide even though it depicts the dependence of P(.1) on sample size for random point samples. In Section 4 it is shown that random points are a conservative estimator of the accuracy for equal-area point patterns.

Likewise, the effect on P(.1) of the climatic frequency of cloud amount can be judged from Figure B-1, subject to recognition that the absolute values of P(.1) in Figure B-1 are for a 10-point random pattern.

5.4.2 Interpretation of Data.

Once taken and relayed to the ground, the data are readily con-

vertible into the 3 operational parameters desired: A) total cloud amount in the target area, B) base and top of each cloud layer present, and C) cloud amount in each layer.

The inference of total cloud amount (within the altitude range sampled) is so simple that it could easily be derived by an onboard counting circuit. The total cloud amount is nothing but the fraction of profiles on which cloudiness was encountered at any level.

Recognizing the individual cloud layers and fixing the base and top of each is probably most easily done subjectively from a simple, side-by-side plot of the profiles.

Once the base and top of a layer have been established, the cloud amount for the layer is merely the average number of cloudy points among all profiles and within the altitude range of the layer. This average is the equivalent of estimating the cloud fraction for all levels sampled within the layer and then averaging the level fractions.

6. CONVERSION FROM CLOUD FRACTION TO CLOUD COVER

Throughout the report to this point we have been concerned with estimating projected cloud fraction whereas the contract Statement of Work asks for estimates of cloud cover. The two measures of cloud amount may differ because the projected fraction, or earth cover, is not sensitive to cloud thickness while the sides of distant cloud elements do contribute to sky cover, the fractional obscuration of the sky when viewed from a point on the ground. Therefore, one would expect: A) the two measures to be identical whenever the earth cover is either 0 or 10 tenths, and B) sky cover to be somewhat the greater for intermediate values of earth cover.

To explore this relationship quantitatively, we exploited an analysis already performed on whole-sky photographs taken in conjunction with standard sky cover observations.[2] These observations were made daily at 0900, 1200 and 1500 CST over a span of more than 3 years, at the National Weather Service observing site in Columbia, MO. In the original analysis, 2,805 photographs and matching observations of sky cover were used to derive frequency distributions of cloud cover by sky cover and by sector of the celestial dome.

Table 3 of the reference gives, as a function of sky cover, the cumulative frequency of cloud-free fraction in a circle of 50° angular diameter centered on the zenith. From this was derived Figure 6.1 showing the average sky cover, N, as a function of N(50), the cloud cover of the 50° sector. Since the viewing angle of this sector departs so slightly from vertical, the cloud cover for the sector should approximately equal the cloud fraction, NA(50), of the sector. Initially, it was our further expectation that, when averaged over enough cases, the cloud fraction, NA, of the total sky and NA(50) should be almost equal. However, Figure 6.1 clearly invalidates this hypothesis, for it would mean that total cloud fraction exceeds total sky cover in the upper range of cloud fraction. A scale phenomenon is responsible for the fact that, even on average, NA(50) and NA are unequal. (Not including the extreme points — N(50) = 0 and N(50) = 10 — which account

2. Lund, I. A., D. D. Grantham and R. E. Davis, 1980: Estimating probabilities of cloud-free fields-of-view from the earth through the atmosphere. Journal of Applied Meteorology, 19: 452-463.

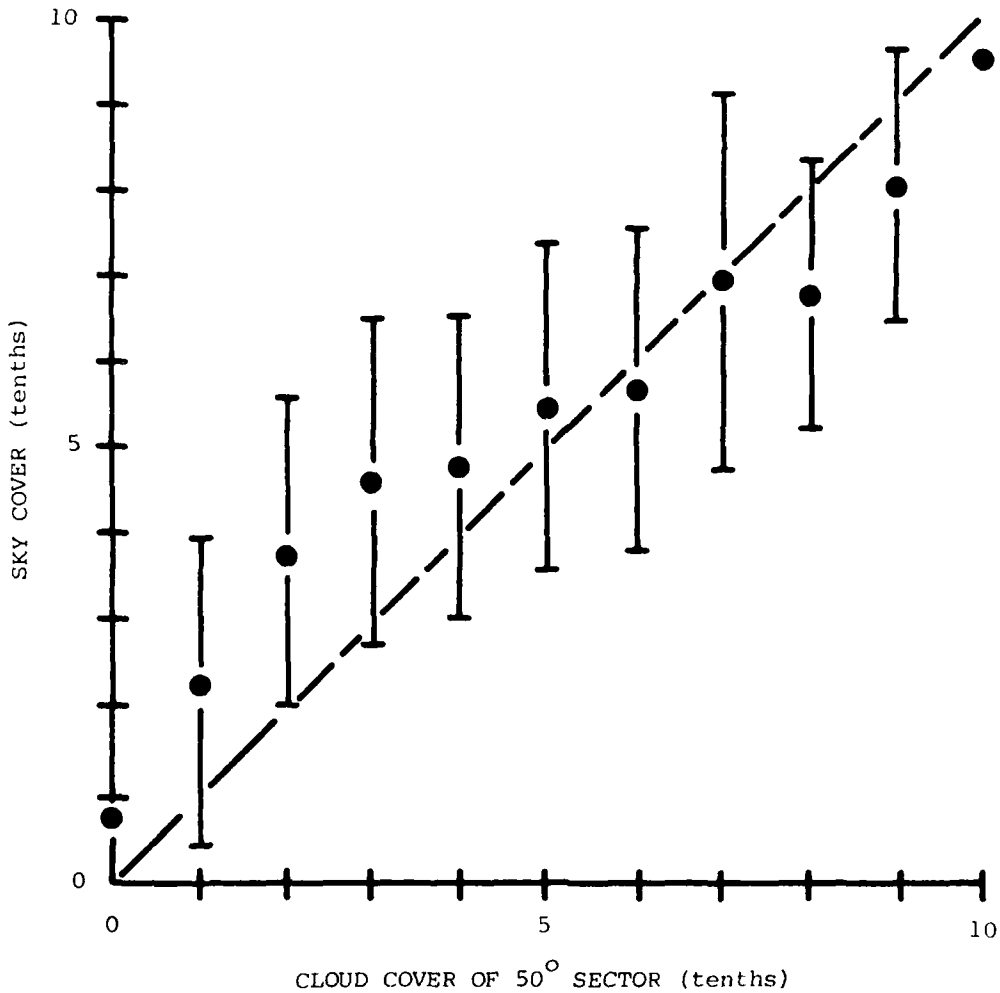


Figure 6.1. Mean Sky Cover and Standard Deviation as a Function of Cloud Cover of the Central 50° Sector.

for 75% of the cases, each point in Figure 6.1 is based on an average of just under 80 cases.)

The existence of this scale phenomenon is most easily recognized in the extreme cases. Consider the cases for which the narrow overhead sector is clear. In at least some of these cases there must be cloudiness elsewhere in the sky. Consequently the average of N for these cases must be greater than 0. At the other extreme, the average of N for all cases in which NA(50) = 1 must be less than 1. The magnitude of the phenomenon probably depends on the size of the sector, which is 10% of the total sky for the 50^{o} sector.

To quantify this scale phenomenon, we return to our own cloud data. Figure 6.2 shows how the cloud fraction, NA, of the 100x100 km total area varies with NA(Q), the cloud fraction of the quadrant. This plot, which is based on the 82 cases of the development sample of cloud fields, confirms the expected shape of the relationship. The line of regression is

$$NA = 1.18 + 0.735\ NA(Q) \qquad (6.1)$$

with a correlation coefficient of 0.99.

Despite the difference in sector size — 25% for the quadrant vs. 10% for the 50^{o} sector — we assume that Eq. (6.1) holds for the dependence of NA on NA(50). This enables Figure 6.1 to be converted into Figure 6.3 depicting sky cover as a function of earth cover — just the relationship that we have been seeking. The ± 1/10 confidence limits for earth cover are also shown, and the sigma-bars in Figure 6.1 apply to the ordinates of the points in Figure 6.3.

Figure 6.3 shows that sky cover tends to exceed earth cover by a small amount, particularly for scattered cloudiness. However, the difference is erratic and is comparable in magnitude to the noise level of the data. Therefore, we feel that in practice there is no basis here for drawing a distinction between sky cover and earth cover. In other words, the projected cloud fraction for the area as derived from the line or point samples discussed in earlier sections should be taken as equal to the corresponding cloud cover.

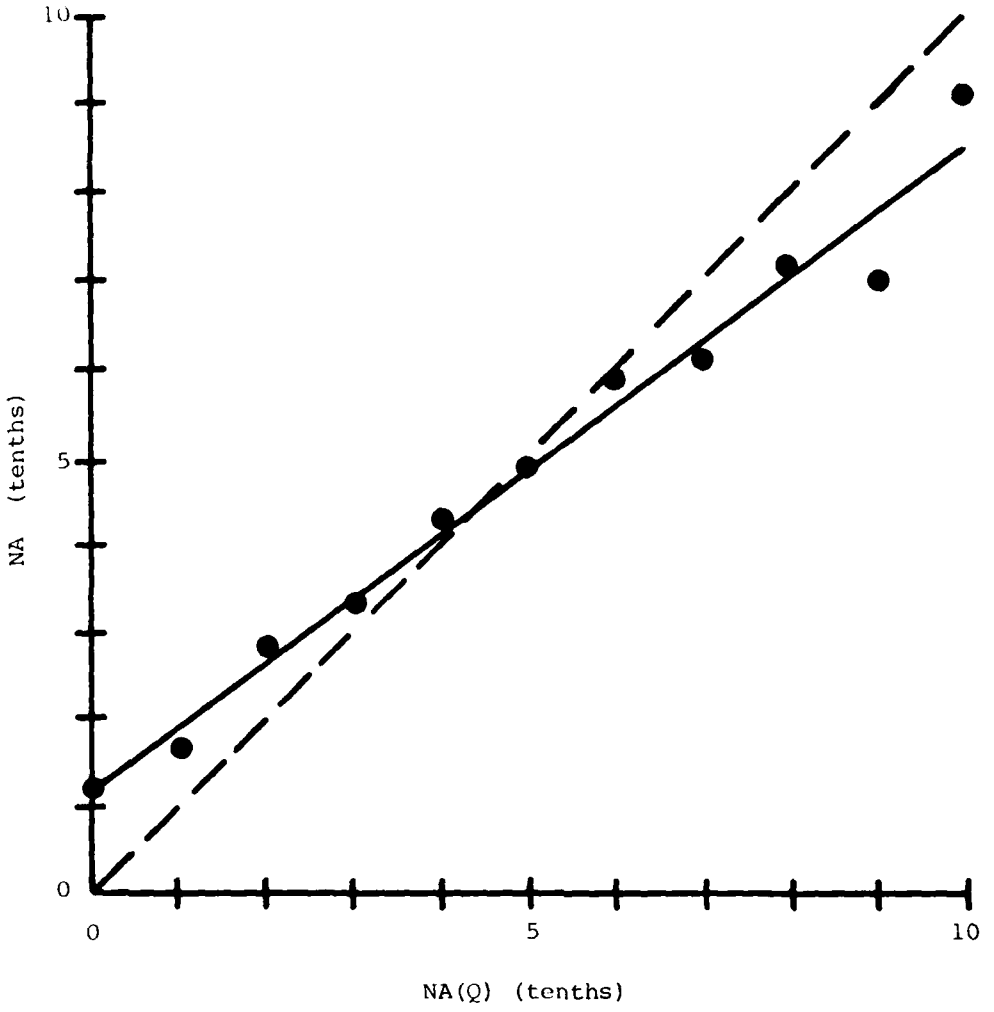


Figure 6.2. Cloud Fraction of Total Sky (NA) as a Function of Cloud Fraction of a Quadrant, NA(Q).

——— Line of Regression.

----- 45^{O} Line.

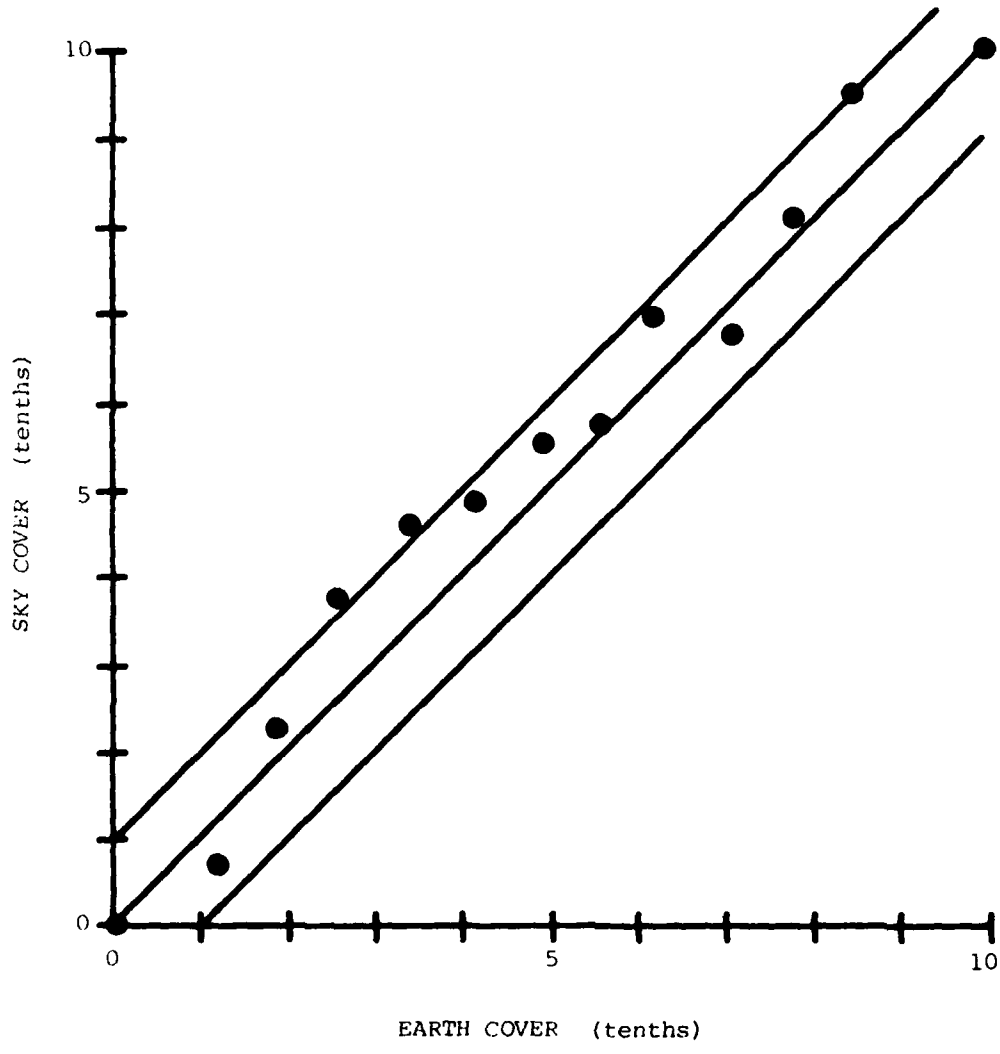


Figure 6.3. Sky Cover as a Function of Earth Cover.
Lines Denote Confidence Limits of Earth Cover.

APPENDIX A

OBSERVED CLOUD FIELDS

McIDAS (Man-Computer Interactive Data Access System) is a minicomputer-based system developed at the University of Wisconsin and designed for the gathering and display of meteorological data. Among its capabilities are the analysis and contouring of conventional meteorological parameters (temperature, pressure, vorticity, streamlines, etc.), the plotting of temperature and moisture soundings, and the depiction of satellite imagery (visible and infrared). Plotting routines are prompted via the terminal keyboard and displayed in color on a CRT.

As an interactive system McIDAS is a very powerful tool that can be put to many varied uses. It is especially well suited for the investigation of clouds since both conventional data and satellite pictures are available for scrutiny.

The cloud data of interest in this report were "half-mile" resolution visible imagery. In particular the goal was to acquire a representative cross section of samples of single layer clouds that covered the spectrum from nearly clear to nearly overcast sky conditions.

Data collection commenced on December 10, 1980 and continued until May 18, 1981. Since some satellite images had been archived on Betamax tapes, it was possible to obtain samples from as far back as May 6, 1980. In all, 132 separate cases were selected.

The usual operating procedure was to examine a satellite photograph of low resolution in the morning in order to identify areas where single layer clouds were located. During the cold weather months care had to be exercised to insure that areas where snowcover was present were not selected for sampling. This precaution was taken to prevent the possibility of ambiguity between reflections from a snow surface and those from a cloud top. Table A-1 shows the locations at which the cloud samples were taken. A large proportion of the cases was taken from the southeastern United States because of the snowcover problem and the fact that, since other users of McIDAS were mainly interested in the eastern half of the country, most satellite pictures

TABLE A-1. LOCATIONS OF SAMPLES.

Location	Number of Samples
Florida	25
Virginia	9
North Carolina	9
Alabama	7
Georgia	6
Pennsylvania	5
Texas	5
Louisiana	4
Colorado	4
Tennessee	3
Mississippi	3
South Carolina	3
Ontario, Canada	3
South Dakota	2
Michigan	2
Ohio	2
New Mexico	2
New Hampshire	2
Minnesota	2
Montana	1
Alberta, Canada	1
Idaho	1
Kentucky	1
West Virginia	1
Northern Mexico	1
Nevada	1
New York	1
Nebraska	1
Maine	1
Quebec, Canada	1
Florida Keys	1
Illinois	1
New Jersey	1
Massachusetts	1
Arkansas	2
Florida coastal waters	7
North Carolina coastal waters	4
Gulf of Mexico	3
New Jersey coastal waters	1
South Carolina coastal waters	1
Lake Ontario	1
TOTAL	132

were centered there.

Once a general area was selected from the low resolution satellite photograph, a higher resolution (nominal half-mile) image was ingested into McIDAS. Since resolution is a function of distance from the sub-satellite point, the actual resolution varied from .620-1.346 miles in the north-south direction to .473-.860 miles in the east-west direction within the cases considered. From the high resolution image an array of 135x135 pixels was picked and this constituted a single sample.

Local noon was often chosen as sampling time so that the sun would be high in the sky and ground-cloud contrast would be a maximum. Each pixel in the satellite image has a brightness (watts cm^{-2} $steradian^{-1}$) associated with it. After adjustment for calibration each pixel was labelled with a brightness count in the range 0-255. Since the purpose of obtaining the satellite data was to produce a field that explicitly depicted the location of cloud elements, a threshold brightness for cloud designation had to be chosen. Any pixel with a brightness above the threshold was designated as cloud. A value of 100 counts was most often chosen as the brightness threshold between cloud and no cloud although the range was from 50-100 counts. Threshold brightness depended on time of day of sampling (sun angle) and type and thickness of cloud (reflectivity). Printed output of x's for clouds and blanks for no clouds was produced for each sample at the time of selection.

There is no assurance, of course, that the subjectively chosen threshold value of brightness precisely defines the cloud boundary. However, since our purposes require only realistic patterns of clouds and spaces, it matters little whether the assumed cloud boundary in a particular instance lies somewhat inside or outside the actual boundary.

Since visual inspection of the satellite image, alone, is not sufficient to guarantee that the clouds were restricted to a single layer, several other sources of information were consulted. These included Circuit "A" reports, synoptic charts and use of McIDAS to overlay surface weather observation on the satellite image. Despite these precautions there still can be no assurance that all the cloud samples were restricted to a single layer.

Once a particular sample was selected, McIDAS wrote the brightness distribution in the area of interest onto magnetic tape. Further processing required use of a large mainframe computer (CDC 6600), for which software was written for data conversion and unpacking. This included a short program which calculated the distance between pixels and then output the number of pixels necessary so that each sample area would be approximately 100 km x 100 km.

Upon examination of the entire cloud field (via the printouts) a suitable sub-area was chosen such that the correct number of pixels for the 100 km^2 area would be analyzed. The decision as to which sub-area to choose was arrived at subjectively. Additional software was developed for the data analysis.

APPENDIX B

THEORETICAL SAMPLING MODEL

Visualize a cloud layer covering some fraction, NA, of the target area. Suppose that some level within the cloud is sampled at a randomly positioned point. The probability that this point is in cloud is NA, the chance of its being in clear air is (1-NA).

Now suppose that the level is sampled at n random points. The probability that all points are in cloud is $(NA)^n$, that all are in clear air $(1-NA)^n$. The probability, P(N), that any intermediate number, N, of the n points is in cloud is given by the binomial distribution:

$$P(N) = \frac{n!}{N!\ (n-N)!}\ (NA)^N (1-NA)^{n-N}. \qquad (B-1)$$

In Table B-1 is shown the binomial distribution of frequencies (probabilities) corresponding to n = 10 and NA = .444. This distribution is labeled F1. The distribution labeled F2, which was transferred from Table 2.1, characterizes the cloud fraction of 60-km passes in one of our observed cloud fields. NA = .444 for F2 also. The comparison is shown merely for general interest. There is no reason to expect close agreement between the two distributions.

TABLE B-1. FREQUENCY (F) OF CLOUD FRACTION (NP) OF A SAMPLE. F1 is the binomial distribution for a set of 10 points. F2, which was taken from Table 2.1, is for 60-km passes. See text for P(.1).

NP (tenths)	0	1	2	3	4	5	6	7	8	9	10	P(.1)
F1 (%)	0	2	8	17	24	23	15	7	2	0	)	64
F2 (%)	1	3	7	11	19	23	18	11	5	2	0	53

As a safety measure, the applicability of the binomial distribution to our cloud data was directly tested and confirmed. Results will be illustrated later in this section.

The 11x11 matrix that forms the main portion of Table B-2A is merely the evaluation of Eq. (B-1) for n = 10 and for all integral 10ths of NA. Each row states the frequency distribution of sample

cloud fraction, NP, corresponding to the particular value of areal cloud fraction, NA. For convenience, the principal diagonal has been lined in.

We define an index of sampling accuracy, denoted P(.1), as the frequency of cases for which sample and areal cloud fraction agree to within 1/10 — i.e., $| NP - NA | \leq 1$. Thus, for each row in the matrix, P(.1) is the sum of the frequency on the diagonal and the two horizontally flanking values, except that there is but one flanking value when NA = 0 or 1. P(.1) is tabulated in the final column of Table B-2A. It is symmetric about NA = 5, where it is at minimum. This illustrates a general truth, namely, that cloud sampling is commonly least accurate when the area is half clouded.

Another way to describe the basic matrix is that it embodies the conditional frequency of sample cloud fraction, given the areal cloud fraction. But the contract Statement of Work poses the converse question: given the sample cloud fraction, what is the frequency distribution of areal fraction? It is possible to transform between the two formulations with the aid of one of Bayes' rules, which will now be derived.

Consider the joint frequency of NA and NP, P(NA,NP) — i.e., the frequency of the simultaneous occurrence of specified values of areal and sample cloud fractions. This can be evaluated as the product of the conditional frequency of NP, given NA, and the unconditional frequency of NA:

$$P(NA,NP) = P(NP|NA) \times P(NA). \qquad (B-2)$$

Conversely, the joint frequency of NP and NA can be expressed as

$$P(NP,NA) = P(NA|NP) \times P(NP). \qquad (B-3)$$

But, by definition, P(NA,NP) = P(NP,NA), and from Eqs. (B-2) and (B-3) it follows that

$$P(NA|NP) = P(NP|NA) \times P(NA)/P(NP). \qquad (B-4)$$

The conditional frequency on the lefthand side of Eq. (B-4) is the answer to the SOW's question. The first factor on the righthand

TABLE B-2 CONDITIONAL FREQUENCY (%) OF CLOUD FRACTION (TENTHS), BASED ON SAMPLE OF 10 RANDOM POINTS.

A. Given the Areal Fraction (NA).
B. Given the Sample Fraction (NP).
See text for other details.

	NA \ NP	0	1	2	3	4	5	6	7	8	9	10	P(NA)	P(.1)
	0	100	0	0	0	0	0	0	0	0	0	0	25	100
	1	35	39	19	6	1	0	0	0	0	0	0	04	93
	2	11	27	30	20	9	3	1	0	0	0	0	03	77
	3	3	12	23	27	20	10	4	1	0	0	0	04	70
	4	1	4	12	21	25	20	11	4	1	0	0	04	67
(A)	5	0	1	4	12	21	25	21	12	4	1	0	01	66
	6	0	0	1	4	11	20	25	21	12	4	1	04	67
	7	0	0	0	1	4	10	20	27	23	12	3	03	70
	8	0	0	0	0	1	3	9	20	30	27	11	04	77
	9	0	0	0	0	0	0	1	6	19	39	35	02	93
	10	0	0	0	0	0	0	0	0	0	0	100	46	100

	NP \ NA	0	1	2	3	4	5	6	7	8	9	10	P(NP)	P(.1)
	0	93	5	1	0	0	0	0	0	0	0	0	27	98
	1	0	51	27	16	5	0	0	0	0	0	0	03	78
	2	0	24	28	29	15	1	1	0	0	0	0	03	82
	3	0	7	20	35	28	4	6	1	0	0	0	03	82
	4	0	2	9	28	35	7	15	4	1	0	0	03	69
(B)	5	0	0	3	15	29	9	29	11	4	0	0	03	67
	6	0	0	1	5	16	7	36	21	13	1	0	03	65
	7	0	0	0	1	6	4	30	28	28	4	0	03	85
	8	0	0	0	0	1	2	17	24	42	13	0	03	80
	9	0	0	0	0	0	0	7	15	45	32	0	02	77
	10	0	0	0	0	0	0	0	0	1	1	97	47	99

TABLE B-3 SAME AS TABLE B-2B, EXCEPT FOR CLOUD CLIMATOLOGY OF FT. RUCKER, AL, IN JULY.

NP \ NA	0	1	2	3	4	5	6	7	8	9	10	P(NP)	P(.1)
0	81	12	5	1	0	0	0	0	0	0	0	13	93
1	0	37	38	17	6	1	0	0	0	0	0	04	75
2	0	16	35	28	15	5	1	0	0	0	0	05	79
3	0	4	22	30	25	12	4	1	0	0	0	06	77
4	0	1	10	22	28	21	11	6	1	0	0	06	71
5	0	0	3	11	22	24	19	16	5	0	0	06	65
6	0	0	1	4	11	18	22	28	15	1	0	06	68
7	0	0	0	1	4	9	17	33	30	7	0	07	80
8	0	0	0	0	1	3	8	26	41	20	0	08	87
9	0	0	0	0	0	1	3	14	39	43	0	08	82
10	0	0	0	0	0	0	0	1	4	9	86	31	95

side is just the conditional frequency in Table B-2A.* If we assume that the sampling takes place randomly in time or on a day-to-day basis, then the unconditional frequency of NA is nothing but the climatic frequency, which is knowable. But what about the unconditional frequency of sample fraction, P(NP)? This can be deduced by recognizing that a sample fraction occurs only in conjunction with some area fraction. Thus, the unconditional frequency of a particular value of NP is the sum of its joint frequency with all possible values on NA. In other words,

$$P(NP) = \sum_{NA} P(NP|NA) \times P(NA) \qquad \text{(B-5)}$$

and, finally,

$$P(NA|NP) = \frac{P(NP|NA) \times P(NA)}{\sum_{NA} P(NP|NA) \times P(NA)} \; . \qquad \text{(B-6)}$$

Note that P(NA|NP) is dependent on the climatic frequency of the areal fraction. This climatic frequency, P(NA), is tabulated in the next-to-last column of Table B-2A. The particular distribution is for cloudiness in the altitude range 0-3 Kft at Fulda, FRG, in January, 1200-1400 LST. These values were derived by Lund using a model developed by Gringorten.[3]

Table B-2B uses this P(NA) to construct from Eq. (B-6) a matrix that is equivalent to the A-matrix except that now the sampled cloud fraction, rather than the area fraction, is the predictor. Again, the last column tallies sampling accuracy, P(.1), as a function of sample cloud fraction.

To show how this second matrix depends on climatology, Table B-3 was evaluated for the areal cloud frequency at Ft. Rucker, AL, in July. This is the B-matrix only, for the A-matrix depends only on the number of points in the sample and is, therefore, the same for the two cases. The climatic frequencies used for Table B-3 are listed in Table B-4.

* As previously noted, because of roundoff, the sum of frequencies is not always precisely 100%.

3. Gringorten, I.I., 1981: Climatic probabilities of the vertical distribution of cloud cover. AFGL-TN (in press).

B-matrices for a 10-point random sample were also constructed for the 8 additional cloud climatologies listed in Table B-4, but they will not be reproduced in detail here. Instead, they will be summarized below.

Although the Gringorten technique was not designed to treat layers of zero-thickness, it can be pushed formally to this limit, which is the "level" of our sampling problem. Consequently, evaluations for a mid-level accompany each layer in Table B-4.

In Table B-4, "Mean" is the average cloudiness. "U" is an index formed by summing the frequencies for NA = 0 and NA = 1. It is a crude measure of the "U-shapedness" of the frequency distribution, and its significance will become evident shortly. "Rank" is the order of the tabulated frequency distributions in terms of the U-index, with 1 assigned to the lowest value of U. Note that, as expectable, the distribution for a level is consistently more U-shaped than that of the embedding layer.

TABLE B-4 CLIMATOLOGICAL FREQUENCY (%) OF AREAL CLOUDINESS.

Ft. Rucker is from the Uniform Summary of Weather Observations.

Fulda and Adana are for January, 1200-1400 LST, and were generated with the Gringorten technique.

	FT. RUCKER		FULDA				ADANA/INCIRLIK			
			0-3	2	6-10	8	0-3	2	6-10	8
NA	JUL	OCT	KFT	KFT	KFT	KFT	KFT	KFT	KFT	KFT
0	10	39	25	85	44	76	86	94	66	70
1	4	7	4	2	4	4	5	2	10	9
2	6	5	3	2	4	1	2	1	5	5
3	6	4	4	1	4	2	2	1	4	3
4	6	4	4	2	3	3	1	1	3	2
5	6	3	1	1	3	1	1	0	2	3
6	6	3	4	1	2	2	1	0	2	1
7	9	4	3	0	4	1	1	1	2	2
8	11	5	4	1	3	1	0	0	2	1
9	9	4	2	1	3	2	0	0	2	2
10	27	24	46	4	26	7	1	0	2	2
MEAN	6	4	6	1	4	1	0	0	1	1
U	37	63	71	89	70	83	87	94	68	72
RANK	1	2	5	9	4	7	8	10	3	6

How P(.1) varies with cloud amount is interesting, but to facilitate comparison of sampling strategies in terms of accuracy, it would be more useful to have some sort of lumped value. However, a simple average of P(.1) stands to be misleading. Because P(.1) is a significant function of cloud fraction, a straight average would not correctly depict the accuracy that could be expected from the particular strategy if it were to be used randomly in time or on a day-to-day basis. For a proper estimate of this expectation, the average of P(.1) must be weighted by the climatological cloud frequency. This means, of course, that the expected sampling accuracy varies with location, season, time of day, altitude, etc., as well as with size of sample.

In Table B-5, "U" is the shape index mentioned earlier, and several averages of sampling accuracy are presented for each of the 10 climatologies in Table B-4. P'(.1) refers to the accuracy index when areal fraction is the predictor. P(.1) is the same measure when the point-sample fraction is the predictor. Without a superscript w, the average is unweighted. Note that the unweighted average of P'(.1) is the same for all 10 cases. This is as it should be, because P'(.1) depends only on the number of points in the random sample, 10 here, and not at all on the climatological cloud frequency. Note that $\overline{P(.1)}$, the accuracy index when the sample fraction is used as a predictor, is slightly lower on average than P'(.1). On the other hand, the weighted mean is invariably and significantly larger than either of the unweighted values. It is this weighted value that is the meaningful estimate of average accuracy expected in practice.

Strictly speaking, the two distributions for Ft. Rucker are inappropriate here because they refer to total sky cover. The other 8 distributions were custom-tailored to represent conditions in a layer or at a level, which is what our scenario calls for. The justification for introducing these "alien" data is that they extend the range of U available for Figure B-1, which shows a clean relation between $\overline{P(.1)}^{w}$ and U. The lowest point corresponds to $\overline{P'(.1)}$, which may be viewed as the weighted mean for a uniform distribution. For such a distribution U = 18%.

In Table B-5 there is no climatologically weighted average for P'(.1). The reason for this is that it is necessarily identical to $\overline{P(.1)}^{w}$. In forming an average, the same matrix elements are summed

whether NA or NP is the predictor — namely all elements for which $|NA-NP| \leq 1$. Basically the values in these positions differ between the A- and B-matrices, but when weighted by the appropriate climatological frequency, what is summed is terms of the form $P(N|A) \times P(A)$ or $P(A|P) \times P(P)$. In either case, the result is $P(A,P)$. In short, the climatologically weighted mean of our index of sampling accuracy is simply the sum of the joint frequency of NA and NP taken over a 3-element-wide band along the principal diagonal.

TABLE B-5 AVERAGED SAMPLING ACCURACY (%) FOR EACH OF 10 CLIMATOLOGICAL CLOUD FREQUENCIES.

Based on binomial distribution for n = 10.
See text for significance of the several values of P(.1)

STATION		U	$\overline{P'(.1)}$	$\overline{P(.1)}$	$\overline{P(.1)}^{W}$
Ft. Rucker	Jul	37	79.9	79.4	84.8
"	Oct	63	"	81.0	91.6
Fulda	0-3 kft	71	"	80.2	92.9
"	2 kft	89	"	80.6	97.5
"	6-10 kft	70	"	80.5	92.8
"	8 kft	83	"	80.0	96.2
Adana/Incirlik	0-3 kft	87	"	71.7	97.3
"	2 kft	94	"	65.1	98.7
"	6-10 kft	68	"	79.5	93.4
"	8 kft	72	"	79.5	94.3
Mean			79.9	77.7	93.9
RMS Δ				5.39	17.9
Mean $\|\Delta\|$				2.71	16.2

As noted earlier, an acid test was conducted to verify that the mean of random point samples from one of our cloud fields is distributed binomially. The test was conducted separately for N = 5, 10, 20, and 30. The procedure will be illustrated here for the case of N = 10.

Ten points are randomly positioned in one of the 100x100 km cloud fields. NOP, the number of points falling within cloud, is counted.

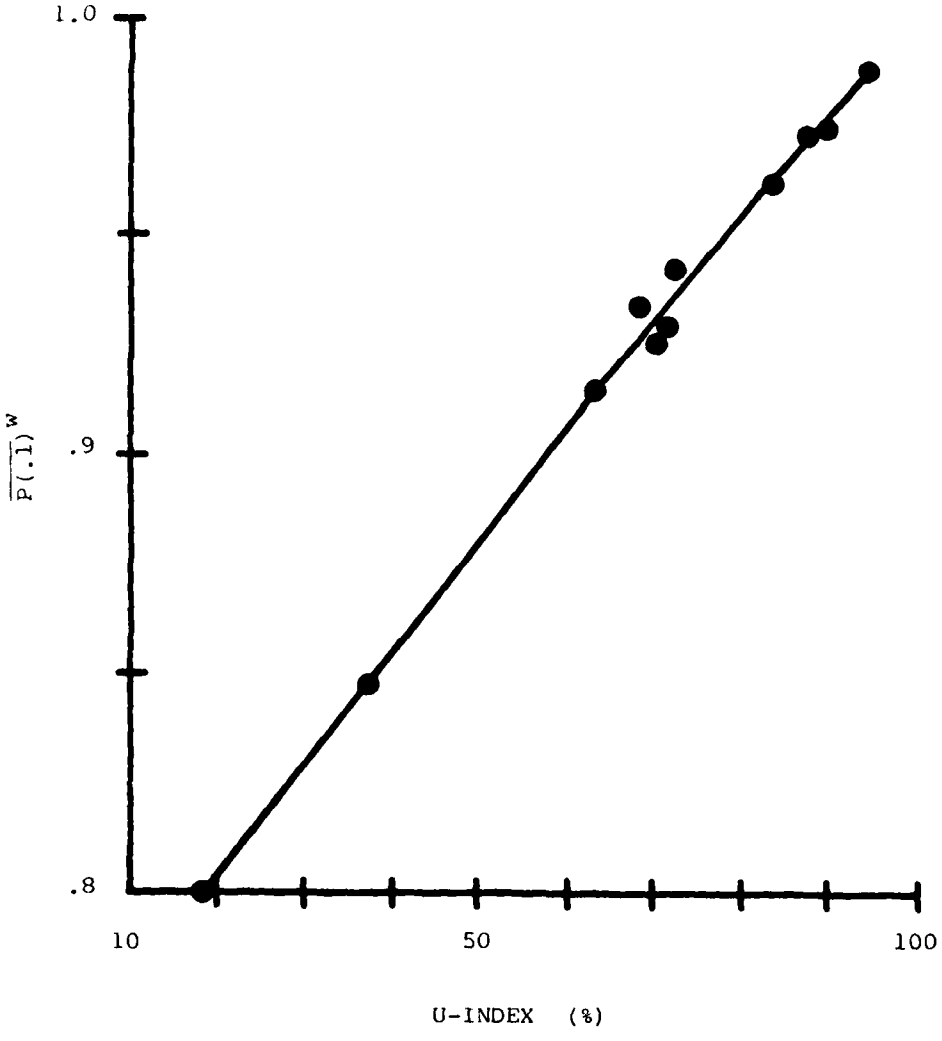


Figure B-1. Weighted Sampling Accuracy as a Function of the U-Index of Cloud Frequency. Based on Binomial Distribution for n = 10.

This constitutes 1 trial. The trial is repeated for a total of 100, and the frequency distribution of NOP is evaluated. This is the experimental distribution for the particular case. To match it, the binomial distribution is derived for the observed value of NA, the areal cloud fraction, and for N = 10.

All 82 cases in the development sample of cloud fields are processed in this manner. The resulting 82 distributions of NOP are grouped by value of NA (rounded to tenths) and averaged. The upshot is an 11x11 matrix of the observed conditional frequency of NOP, given NA. (In the general case this matrix is 11 x (N+1).) A matrix of theoretical frequencies is generated by performing the same operations on the binomial distributions for the 82 cases. The sampling accuracy, P(.1), is evaluated for each row in both matrices.

The differences between corresponding elements of the two matrices are shown in Table B-6. The close agreement is immediately evident, considering that the row-sum for both matrices is 100 (within roundoff). For more detailed inspection, the theoretical matrix here is almost identical to that shown in Table B-2A, the difference being that the rows here are averages of distributions based on generally non-integral values of NA.

The last column in Table B-6 shows that, not suprisingly, the agreement in P(.1) between observed and theoretical distributions is also very good.

TABLE B-6 DIFFERENCE BETWEEN FREQUENCIES OF OBSERVED RANDOM SAMPLES AND BINOMIAL DISTRIBUTION FOR N = 10.
NA and NP are in tenths; all other values are in %.

NA \ NP	0	1	2	3	4	5	6	7	8	9	10	P(.1)
0	1	0	0	0	0	0	0	0	0	0	0	1
1	0	-1	0	0	0	0	0	0	0	0	0	-1
2	1	-2	2	-1	1	1	0	0	0	0	0	-1
3	-1	1	3	-1	-1	0	-1	0	0	0	0	1
4	-1	-2	-1	0	1	0	0	0	1	0	0	1
5	0	0	0	1	-2	0	0	0	1	0	0	-2
6	0	0	0	0	0	1	0	1	0	0	0	1
7	0	0	0	0	0	-1	0	-3	1	1	0	-2
8	0	0	0	0	0	1	1	-3	-2	2	-1	-3
9	0	0	0	0	0	0	0	-1	2	-1	1	2
10	0	0	0	0	0	0	0	0	0	1	-1	0

Similarly, we found excellent agreement between observed and theoretical frequencies for the other values of N tested: 5, 20, 30. Indeed, the agreement was even better for the larger values of N. Consequently, we conclude without reservation that random point samples taken within cloud fields do follow the binomial distribution.

APPENDIX C

ESTIMATING THE INDEPENDENCE FRACTION

If the cloud sensor is sampled at 1 Hz, a pass across the target area produces more than a thousand points at intervals of less than 50 meters. How good a predictor of the areal cloud cover is the cloud fraction of these points? As outlined in Appendix B, this question could be readily answered in terms of the binomial distribution if the points were independent, but samples so closely spaced are highly correlated. This interdependence inflates the variance of the point fraction and reduces the accuracy of the pass mean cloudiness as an index of area cloud cover. In terms of this variance, the pass behaves as though composed of some smaller number of independent points. Using the procedure described below, for each of the 82 basic cloud fields we evaluated this reduced number, N', and the "Independence Fraction," IF, defined as N'/N where N is the actual number of points in the sample.

For each case the variance, σ_N^2, of the half-row cloud fraction was directly evaluated for the 100x100 km area. On the average, these fractions were based on 62 points separated by .82 km, and σ_N^2 was based on 166 values.

Since the point value is either 0 or 1, the mean and mean square of any collective of points are identical. Consequently, considering the 10,000-plus points in the 100x100 km cloud field as the entire population, their variance is NA(1-NA) where NA is the cloud fraction for the whole area which is, of course, the population mean. It is known that the variance, $\sigma_{N'}^2$, of the mean of any subset consisting of N' independent samples is 1/N' times the point variance. We now ask in each case: what value of N' yields the same variance as that observed in the half-rows? This generates the following formulation for the Independence Fraction:

$$IF = N'/N = \frac{NA(1-NA)}{N \times \sigma_N^2}. \qquad (C\text{-}1)$$

Using the half-rows and Eq. (C-1), we evaluated IF for each of the 82 cases. The average value of IF was 0.090 with a standard deviation of 0.054. In other words, in terms of their variance, the

half-row cloud fractions behaved as though based on fewer than 6 points (.090 x 62 = 5.8).

An analogous concept is that of "Independence Unit of Length," IU, which is the average separation of the hypothetical N independent samples. Since the half-rows are 50 km long in all of our cases, IU = 50/N' km. For all 82 cases the average of IU is 12.33 km with a standard deviation of 6.7 km.

Normally, the procedure above cannot be used in practice for want of a sufficient number of passes from which to compute the variance of the pass mean. Hence we derived and tested several formulations for approximating IF from single-pass data. All of these entailed evaluating autocorrelation functions, and all resulted in values of IF that were, on average, larger than the directly derived values discussed above.

Since several of these approximations of IF assume that the sequence of samples forms a Markov chain, we tested this assumption by evaluating the "Markov multiplier" for lags 2-6. The data used were autocorrelations for ℓ = 1-6, based on entire rows, which average 124 points in length. For each lag the autocorrelation coefficient was averaged for the approximately 83 rows in each case. The Markov multiplier is defined as $M_\ell = \rho_\ell / \rho_1^{\ell}$ where ℓ is the lag. These multiples were then averaged over the 82 cases. The results are plotted in Figure C-1. For a Markov sequence, $M_\ell = 1$ for all values of ℓ. Figure C-1 shows that in our cloud samples, the autocorrelation function decays more slowly than Markovian.

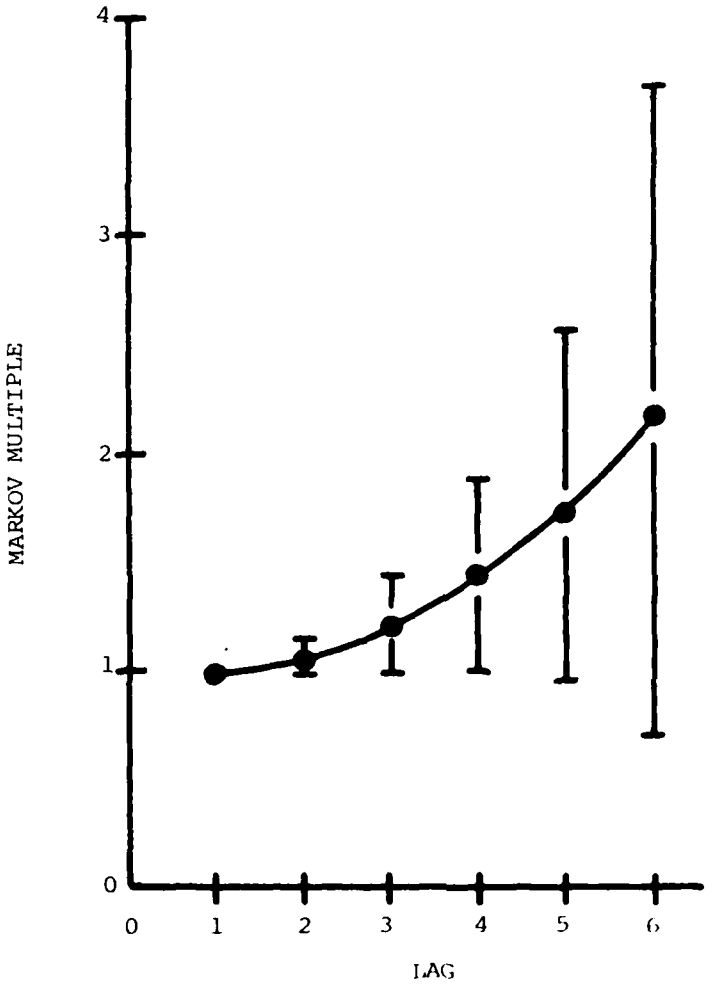


Figure C-1. Average Markov Multiple as a Function of Lag.

APPENDIX D

SOME REFLECTIONS ON HOMOGENEITY

The contract Statement of Work invited us, first, to work the overall problem assuming horizontal homogeneity of the cloud field and, then, to consider and evaluate the effects of non-homogeneity. As noted in the overview, we found it from the outset unnecessary to limit our purview to homogeneous cloud fields. The theoretical approach made no stipulation about the cloud field, only that the sampling points be randomly deployed. The experimental approach was based on cloud fields as nature served them up. Nevertheless, we have given some thought to the pleasures of sampling a homogeneous field.

What is homogeneity? The Statement of Work gives no clue. As a realization of a horizontally homogeneous cloud field, one might well visualize a field of fair-weather cumulus, of size and spacing that are variable but not too much so, and without any mesoscale structure. Any sampling pass made through such a field — provided that it is sufficiently long relative to the "scale size" of the field — should, in a statistical sense, be equivalent to any other pass. The most primitive statistic is the mean cloudiness along the path. Thus, we are led to the definition that a cloud field is homogeneous if and only if all sampling passes of sufficient length yield the same value of cloud fraction. Under such conditions, the sampling problem is trivial. A single horizontal pass of sufficient length yields a flawless estimate of cloud fraction for the area.

While all horizontally homogeneous cloud fields are, thus, equally easy to sample, it does not follow that all inhomogeneous fields are equally difficult to sample. There is one, not-uncommon, situation that is particularly troublesome. The field we describe is the one we call the "cloud front" field, which is shown in Figure D-1. Given an area of interest, which we take to be a square, we break it into exactly two pieces, one overcast and one clear. For further simplification, we assume that each of the two pieces is a rectangle, the "cloud front" being the line between them.

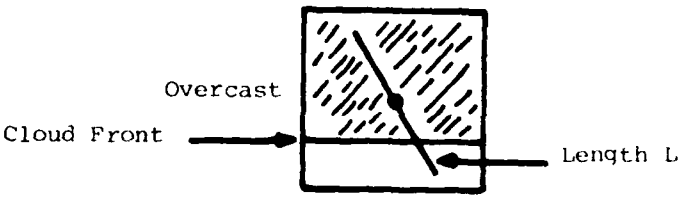


Figure D-1. Geometry of the Cloud Front Sample.

We now perform straight-line sampling of this cloud field. We shall assume that:

(1) the square is of length 2;

(2) the sample line of fixed length L (≤ 2) is centered in the square, and takes no angular preference;

(3) without loss of generality, the areal cloud cover NA is $\geq \frac{1}{2}$.

Let the fraction of our straight-line sample that is in cloud be NL. Some consideration shows that in all cases $NA \leq NL$! (If $NA \leq \frac{1}{2}$ then $NA \geq NL$ always.) The question to be answered is this: with all areal cloud covers NA being equally likely, find the probability P(.1) that NL is within .1 of NA.

A short calculation shows that

$$\Pr\left(|NA-NL| \leq .1 \mid NA\right) = \frac{2}{\pi} \cos^{-1}\left(\min\left\{1, \frac{2NA-1}{L(NA-.4)}\right\}\right). \qquad (D\text{-}1)$$

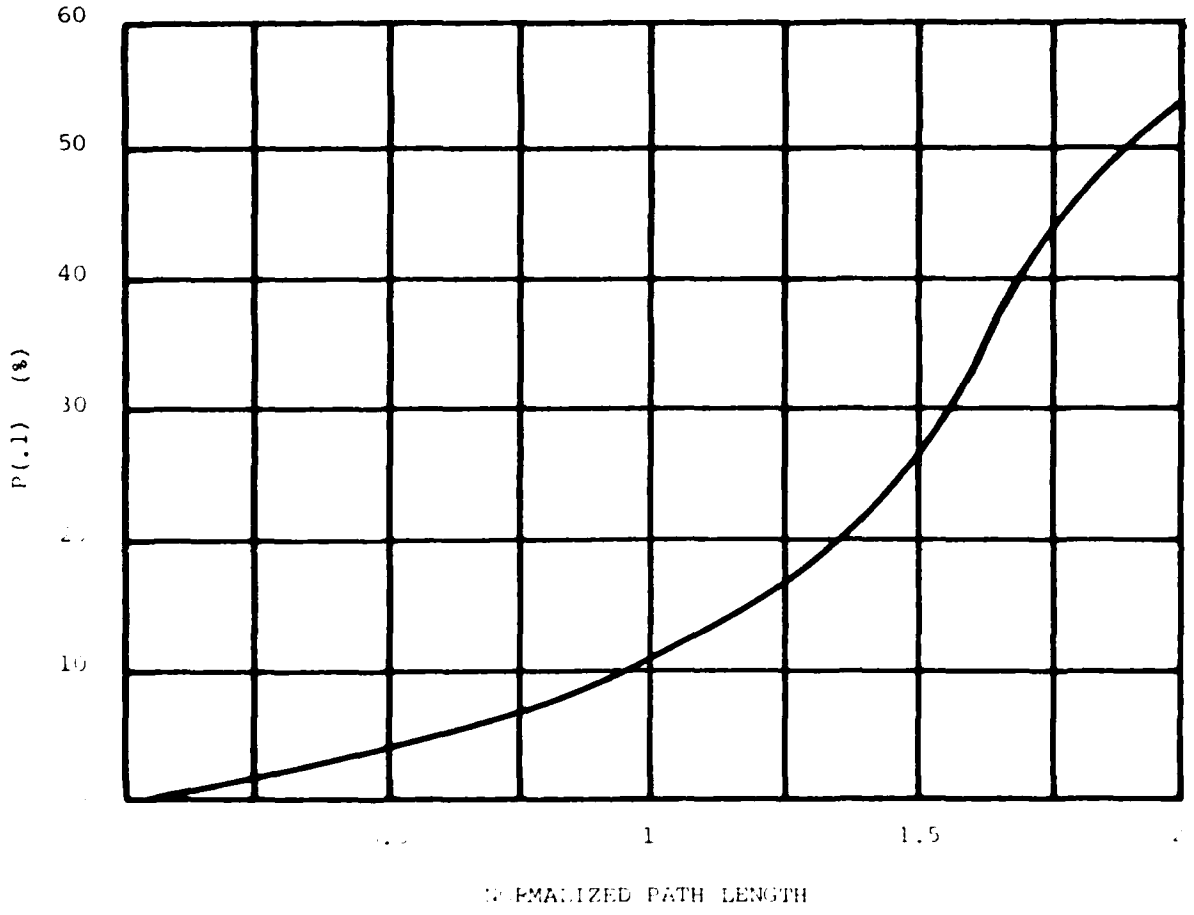
We can thus write

$$P(.1) = \int \Pr(|NA - NL| \leq .1 \mid NA)\, d(NA), \text{ and}$$

$$P(.1) = \frac{2}{\pi} \int_{\frac{1}{2}}^{1} \cos^{-1}\left(\min\left\{1, \frac{2x-1}{L(x-.4)}\right\}\right) dx. \qquad (D\text{-}2)$$

We expect that as L increases from 0 to 2, P(.1) increases as well. We have evaluated Eq. (D-2) for a succession of path lengths L, obtaining the graph shown in Figure D-2.

The "cloud front" field is probably the most extreme example of an inhomogeneous field; yet it certainly occurs with no small probability. If such a field is considered likely on a given day, Figure D-2 shows that even with a straight line pass equal in length to the scale of the region in question (L = 2), the probability of obtaining an accurate sample is just over 50%. We note finally that if instead of sampling along a straight line an equal-area point sample is taken, the resulting P(.1) is larger and is equal to 60% for the 9-point sample, 80% for the 16-point sample.


60
50
40
30
10
P(.1) (%)
1
1.5
NORMALIZED PATH LENGTH

APPENDIX E

A NOVEL METHOD FOR EVALUATING THE ONE-LAG AUTOCORRELATION

In view of the simple, dichotomous nature of cloud observations it is possible to derive the one-lag autocorrelation in an extremely simple form; namely,

$$\rho_1 = 1 - g/nNL(1 - NL) \tag{E-1}$$

where g is the number of separate cloud groups in the series of observations of zeros (clear) and ones (cloud). Therefore g is the number of discrete strings of ones. The remaining terms have their usual meaning: n is the total number of observations in the series and NL is the fraction which are ones.

Eq. (E-1) can be derived by means of two separate approaches. One makes use of probability concepts and the other consists essentially of direct substitution into the defining expression for the one-lag autocorrelation. Both approaches will be illustrated since they each offer somewhat different insights into the nature of Eq. (E-1).

We have a series of uniformly spaced observations along a straight, horizontal line whose elements consist of zero (clear) or one (cloud).

Let $P(0|1)$ be the conditional probability of a zero given a one as the antecedent observation. Let $P(1|1)$ be the conditional probability of a one given a one as the antecedent observation. Then it is apparent that

$$P(1|1) = 1 - P(0|1). \tag{E-2}$$

$P(1|1)$ times the number of ones in the series is merely the sum of the one-lagged products. That is,

$$n\ NL\ P(1|1) = \sum y_i y_{i+1} \tag{E-3}$$

where y_i is the i^{th} observation. Thus,

$$NL\ P(1|1) = \overline{y_i y_{i+1}} \tag{E-4}$$

where the operation ($\overline{\ }$) indicates an average over the domain. From the definition of the one-lag autocorrelation, ρ_1,

$$\rho_1 = (\overline{y_i y_{i+1}} - \bar{y}_i \bar{y}_{i+1}) / \sigma_{y_i} \sigma_{y_{i+1}} \qquad \text{(E-5)}$$

where $\sigma^2_{y_i} = \sigma^2_{y_{i+1}} = \sigma^2_y$ is the variance of the set of observations, we obtain $\overline{y_i y_{i+1}} = \rho_1 \sigma^2_y + \bar{y}^2 = \rho_1 (\overline{y^2} - \bar{y}^2) + \bar{y}^2$

$$= \rho_1 - (\bar{y} - \bar{y}^2) + \bar{y}^2 = \rho_1 (NL - [NL]^2) + (NL)^2.$$

Therefore, from Eq. (E-4)

$$P(1|1) = (1 - NL)\rho_1 + NL = \rho_1 + NL\ (1 - \rho_1). \qquad \text{(E-6)}$$

However, $P(0|1)$ may be expressed as

$$P(0|1) = \sum_x f(x) \ / \ \sum_x xf(x) \qquad \text{(E-7)}$$

where $f(x)$ is the frequency of a string of x ones. Therefore $\sum_x f(x)$ is the number of separate cloud groups (q), and $\sum_x xf(x) = n\ NL$.

The latter quantity is of course the total number of ones in the series of observations. Thus we have,

$$P(0|1) = q/n\ NL \qquad \text{(E-8)}$$

but from Eqs. (E-2) and (E-6) we have

$$P(0|1) = 1 - \rho_1 - NL(1 - \rho_1) = (1 - \rho_1)\ (1 - NL).$$

Therefore, we have

$$\rho_1 = 1 - q/n\ NL(1 - NL). \qquad \text{(E-1)}$$

The above expression for the one-lag autocorrelation can be derived directly from the standard definition of ρ_1, Eq. (E-5). If we are careful to define the domain of the operator ($\overline{\ }$) so as to account for end effects, then the domain of i in Eq. (E-5) is i = 1, 2, ..., n-1.

In terms of quantities which have already been defined,

$$\overline{y_i y_{i+1}} \simeq (nNL - q)/n \qquad \text{(E-9)}$$

$$\bar{y}_i \simeq \bar{y}_{i+1} \simeq NL \qquad \text{(E-10)}$$

$$\sigma^2_{y_{i+1}} \simeq \sigma^2_{y_i} \simeq \overline{y_i^2} - \overline{y_i}^2 \simeq NL - (NL)^2. \qquad \text{(E-11)}$$

Substituting Eq. (E-9), (E-10) and (E-11) into Eq. (E-5) we obtain Eq. (E-1) directly,

$$\rho_1 = \left[(nNL - q) / n - (NL)^2\right] / \left[NL - (NL)^2\right] = 1 - q/nNL\ (1 - NL).$$

In Eqs. (E-9), (E-10), and (E-11) the expressions are indicated as approximate. The approximation arises because the domain of the $\overline{(\quad)}$ operator is i = 1, 2, ..., n-1, whereas in Eq. (E-9) we have divided by n. Similarly, in Eqs. (E-10) and (E-11) NL is the cloud fraction for the entire series of n observations, whereas $\bar{y}_i$ and $\overline{y_{i+1}}$ are averages over only (n-1) observations. The approximation however is certainly very good for all but extremely short series of observations.

Therefore, in any series, where the elements may be expressed as zeros or ones, the simple relation for the one-lag autocorrelation, Eq. (E-1), may find useful application.

APPENDIX F

ADDITIONAL THEORETICAL SAMPLING DISTRIBUTIONS

Here we present some simple sampling models examined during the course of this contract. These models have assisted our understanding and may possibly be of value in future related studies.

Because many atmospheric processes are known to be Markovian in nature we early on made a simple Markov model and studied its characteristics. While the empirical data we later collected through McIDAS appear to be samples of a sub-Markovian process, nevertheless the Markov model is a reasonable one and does provide insight. We proceed to outline its structure.

Let $\{X_n\}$, n = 1, 2, 3, ..., represent the sequence of zeros and ones returned by a straight line pass of the APV. If the areal cloud fraction is NA then for any i the probability that $X_i = 1$ is NA and the probability that X_i is 0 is 1 - NA. We wish to introduce the notion of persistence into our model. Assume that the first N values of the sequence are known. We model the conditional probabilities of X_{n+1} as follows, introducing persistence through the parameter α, with $0 \leq \alpha \leq 1$:

$$\begin{aligned}
&\Pr\ (X_{N+1} = 1 \mid X_N = 1) = \alpha + (1-\alpha)NA \\
&\Pr\ (X_{N+1} = 0 \mid X_N = 1) = (1-\alpha)(1-NA) \\
&\Pr\ (X_{N+1} = 1 \mid X_N = 0) = (1-\alpha)NA \\
&\Pr\ (X_{N+1} = 0 \mid X_n = 0) = \alpha + (1-\alpha)(1-NA).
\end{aligned} \tag{F-1}$$

It is possible to think of sequences whose $(N+1)^{\underline{st}}$ values are dependent upon all of the values $\{X_n\}$, n = 1, 2, ..., N. The fact that the probability law for X_{N+1} in our model Eq. (F-1) depends only on the last value X_N makes our process Markovian. It should be clear that once the value for X_1 is given, Eq. (F-1) inductively determines the probability law for X_2, X_3, X_4, ... To start the process it is natural to take $\Pr\ (X_1 = 1) = NA$ and $\Pr\ (X_1 = 0) = 1 - NA$.

We have now completely specified our (Markovian) model Eq. (F-1). It is not hard to show that for any value of N, the expected value of X_N is NA. Further, the one-lag autocorrelation of the process is α. Because the one-lag autocorrelation is called ρ in earlier sections we shall here make the identification $\alpha = \rho$. It is not hard to see that if $\alpha = 0$ then the X_N's are completely independent. As α increases toward 1 the effect of X_N on X_{N+1} increases, until in the extreme case $\alpha = 1$ all of the X_N's are equal to the value of X_1. It is for this reason that α is considered to be a measure of persistence.

In operational practice a sample sequence of zeros and ones is provided by the ADV, and we wish to draw inferences from it. Before we can do so, however, we need to know more about the model Eq. (F-1); in particular we ask first: given that we know both NA and ρ ($=\alpha$) and that we make a total of N observations, what is the probability that exactly k of our observations are equal to one? Let us write this probability as ${}_NS(k; NA,\rho)$. We have written a program called PROB which calculates S. It is worthwhile to note that

$$\sum_{K=0}^{N} {}_NS(k; NA, \rho) = 1 \qquad \text{(F-2)}$$

$${}_NS(k; NA, 0) = \binom{N}{k} NA^k (1-NA)^{N-k}. \qquad \text{(F-3)}$$

Eq. (F-3) is a statement of the fact that in the limiting case $\rho = 0$ our model reduces to a sequence of N independent trials of the binomial distribution with parameter NA.

For the general case $\rho \neq 0$ it is clear that persistence makes our sample of N observations dependent on one another. We ask as in previous sections: find an appropriate measure of the amount of independent information contained in any sample sequence. We make this question more explicit: using the chi-square test, find the value of N' so that ${}_{N'}S(k; NA, 0)$ best fits ${}_NS(k; NA, \rho)$.

The following table gives the results for the case N = 100; for each (NA, ρ) pair, the entry is the value of N':

TABLE F-1. N' AS A FUNCTION OF NA AND ρ (N=100).

NA	.1	.2	.3	.4	.5	.6	.7	.8	.9
ρ									
.1									
.2			>50			>50			>50
.3									
.4	44	43	43	43	43	43	43	43½	44
.5	35	34	34	34	34	34	34	34	35
.6	25	26	25	25	25	25	26	26	25
.7	15	18	18	18	18	18	18	18	15
.8	14	10	11	12	12	12	12	10	14
.9	17½	8	5	4	4	4	4	8	17½

We now turn the problem around; instead of a probability law being given, a sample of N observations is considered given. In addition, the climatology function CP(ℓ), ℓ = 0, 1, 2, ..., 10 = the climatological probability of the cloud cover being $^{\ell}/_{10}$ is assumed to be given. From the sample we can obtain an N' in various ways, including, for example, using the above table. As in Appendix B, Bayes' Theorem then provides the conditional probabilities Pr (NA|NL) = the probability that the actual cloud cover is NA given the observed cloud fraction NL. The computer program NPRIME performs this calculation for a given climatology, N', and observation, NL. From this output confidence levels may be obtained.

There is another (related) way to use Bayes' Theorem without calculating an N'. Given a sample of N total observations, calculate the observed cloud fraction NL and RHOSAM, the sample one-lag autocorrelation. Just as Bayes' Theorem can be used to work from $\left\{ {}_{N'}S(k;\ NA,\ 0) \right\}$, NA = 0, .1, ..., 1.0 together with a climatology, it can also be used to work from $\left\{ {}_{N}S(k;\ NA,\ RHOSAM) \right\}$, NA = 0, .1, ..., 1.0 and a climatology, calculating as above Pr (NA|NL). Program DPROB performs the calculation.

The final section of this appendix addresses the question of how to proceed to obtain confidence levels from a sample sequence when no climatology is available. One method is to apply Gauss' maximum likelihood estimator to estimate unknown parameters.[4]

4. Larson, H., 1969: Introduction to Probability Theory and Statistical Inference. John Wiley & Sons, pp 253.

Assume that we are given a sample sequence $\{X_n\}$ of N observations. Proceeding as before, we can reduce to N' independent Bernoulli trials in a variety of ways. If N' is large enough, it can be shown that if we want to find P_1 and P_2 such that

$$\Pr(P_1 \leq NA \leq P_2) = 1 - \beta, \text{ where NA is the} \qquad \text{(F-4)}$$

(unknown) cloud cover, then

$$P_1 = NL - \frac{Z}{\sqrt{N'}} \sqrt{NL\ (1 - NL)} \quad \text{and} \qquad \text{(F-5)}$$

$$P_2 = NL + \frac{Z}{\sqrt{N'}} \sqrt{NL\ (1 - NL)}, \text{ where } Z_{1-\beta/2} \qquad \text{(F-6)}$$

is the 100 $(1-\beta/2)$ percentile of the standard (mean 0, variance 1) normal distribution function and NL is the observed cloud fraction of the given sample.

Our last technique combines the notion of a Markov process with Gauss' maximum likelihood estimator.

Assume that we are given a sample $\{X_n\}$ of N observations of some unknown Markov process. Can we find that Markov process which is most likely to have generated our sample set of observations? We have done so, stating the result below.

Given X_n, let

$$\begin{aligned} C_{00} &= \#\text{ of times a 0 follows a 0.} \\ C_{01} &= \#\text{ of times a 1 follows a 0.} \\ C_{10} &= \#\text{ of times a 0 follows a 1.} \\ C_{11} &= \#\text{ of times a 1 follows a 1.} \end{aligned} \qquad \text{(F-7)}$$

Let $$A = \frac{C_{10}}{C_{11}} \quad \text{and} \quad B = \frac{C_{01}}{C_{00}}. \qquad \text{(F-8)}$$

Define

CL $\equiv$ most likely NA,

RHOL $\equiv$ most likely ρ. (F-9)

A calculation shows that

$$CL = B - \frac{1 + A}{A + B + 2AB}, \tag{F-10}$$

$$RHOL = \frac{1 - AB}{(1 + A)(1 + B)}. \tag{F-11}$$

Given the sample $\{X_n\}$ of N observations, one can write in analytical form the probability function $P = P(NA,\rho)$ that is the probability of obtaining the sequence $\{X_n\}$ from the Markov process (NA,ρ). One then finds through elementary calculus that P is maximized at (CL, RHOL), yielding the above result. It is not unreasonable to expect that levels of confidence for this result can be obtained from the function P. We have thus outlined a scheme whereby under the Markov assumption and with no climatology one can pass from a line sample to the most likely generating Markov process. Work that might be done in the future includes finding the confidence interval for this result.


DATE
FILME
4-8